



**Universidade de
Aveiro**

2015

Departamento de Electrónica, Telecomunicações e
Informática

**TIAGO LOPES FERREIRA
BRITO**

**Visualização de Fluxos Migratórios: Os Candidatos ao Ensino
Superior Público (2012 a 2014)**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Sistemas de Informação, realizada sob a orientação científica do Dr.-Ing. Joaquim João Estrela Ribeiro Silvestre Madeira, Professor Auxiliar do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro, e da Doutora Maria Beatriz Alves de Sousa Santos, Professora Associada com Agregação do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro.

o júri

presidente

Professor Doutor José Manuel Matos Moreira

Professor Auxiliar da Universidade de Aveiro

arguente principal

Professora Doutora Maria Beatriz Duarte Pereira do Carmo

Professora Auxiliar da Faculdade de Ciências da Universidade de Lisboa

orientador

Professor Dr.-Ing. Joaquim João Estrela Ribeiro Silvestre Madeira

Professor Auxiliar da Universidade de Aveiro

agradecimentos

A todos que direta ou indiretamente contribuíram para a elaboração deste trabalho. Em especial à Direção-Geral do Ensino Superior (DGES) pela disponibilização dos dados utilizados como caso de estudo nesta dissertação.

palavras-chave

visualização de informação, fluxos migratórios, ensino superior.

resumo

A importância e quantidade da informação criada pelos vários sectores da sociedade, tornam a visualização e compreensão numa das tarefas mais importantes de hoje em dia. A utilização da Visualização de Informação para representar diferentes tipos de dados de uma forma mais compreensível tem-se generalizado em várias áreas. Esta dissertação tem como objectivo investigar como é que a Visualização de Informação poderá ser usada para aumentar a legibilidade e usabilidade de dados de migração dos candidatos ao Ensino Superior português. Atualmente existem um conjunto diverso de visualizações para ilustrar este tipo de informação. Este trabalho pretende desenvolver visualizações eficientes e interativas para analisar as migrações dos candidatos ao Ensino Superior, utilizando apenas ferramentas Web e validar se os utilizador conseguem compreender a informação apresentada.

Esta dissertação apresenta uma breve introdução à Visualização de Informação, sintetiza as representações de fluxos migratórios mais relevantes até ao momento, apresenta algumas ferramentas e bibliotecas para a sua implementação em ambiente Web. As três representações criadas são apresentadas e validadas através de avaliação heurísticas e testes de usabilidade, exemplificando assim todo o ciclo de desenvolvimento necessário para criar uma conjunto de visualizações utilizando um caso de estudo real.

keywords

information visualization, migration flows, higher education.

abstract

The use of Information Visualization methods to represent different types of data in a more understandable way has become widespread in various fields. This thesis aims to investigate how Information Visualization can be used to increase the readability and usability of the annual data associated with the placement of Portuguese Higher Education candidates. Currently there are multiple visualizations available to illustrate this type of information. The aim is to develop efficient and interactive visualizations to analyze the candidates' placement/migration, using only Web tools, and to validate that the user can effectively use the visualizations and understand the information presented.

This work presents a brief introduction to Information Visualization, summarizes the most important representations for migratory flows, and presents some tools and libraries for an implementation in a Web environment. The three representations developed are presented and validated, through heuristic evaluation and usability testing, thus illustrating the entire development cycle for creating a set of visualizations within a real case study scenario.

Índice

1	INTRODUÇÃO	1
1.1	MOTIVAÇÃO	1
1.2	OBJETIVOS	2
1.3	VISÃO GERAL	2
2	ESTADO DA ARTE	5
2.1	VISUALIZAÇÃO DE INFORMAÇÃO	5
2.1.1	<i>Visualização de Informação versus Visualização de Dados</i>	<i>7</i>
2.2	INTRODUÇÃO AOS FLUXOS MIGRATÓRIOS E DADOS ORIGEM-DESTINO	10
2.3	VISUALIZAÇÃO DE FLUXOS MIGRATÓRIOS	12
2.3.1	<i>Evolução.....</i>	<i>12</i>
2.4	MÉTODOS DE VISUALIZAÇÃO DE FLUXOS MIGRATÓRIOS.....	16
2.4.1	<i>Fluxos georreferenciados.....</i>	<i>16</i>
2.4.2	<i>Grafo.....</i>	<i>17</i>
2.4.3	<i>Representação Matricial.....</i>	<i>18</i>
2.4.4	<i>Diagrama de Sankey.....</i>	<i>18</i>
2.4.5	<i>Diagrama cordal.....</i>	<i>20</i>
2.4.6	<i>Escalas</i>	<i>21</i>
2.4.7	<i>Utilização de cor</i>	<i>22</i>
2.4.8	<i>Cartografia.....</i>	<i>22</i>
2.5	FERRAMENTAS.....	23
2.5.1	<i>D3.js</i>	<i>24</i>
2.6	CONCLUSÕES.....	26
3	VISUALIZAÇÃO DAS CANDIDATURAS AO ENSINO SUPERIOR PORTUGUÊS.....	29
3.1	CONJUNTO DE DADOS	29

3.1.1	<i>Dados dos Concursos Nacionais de Acesso (1ª fase, 2012 a 2014)</i>	30
3.1.2	<i>Pré-Processamento dos dados</i>	33
3.1.3	<i>Definições gerais</i>	36
3.2	REPRESENTAÇÕES	37
3.2.1	<i>Filtros e controles</i>	38
3.2.2	<i>Matriz de adjacências</i>	40
3.2.3	<i>Diagrama cordal</i>	43
3.2.4	<i>Georreferenciação de fluxos</i>	47
3.2.5	<i>Ferramentas complementares</i>	55
3.2.6	<i>Resultados da exploração interativa dados</i>	57
4	AVALIAÇÃO	59
4.1	AVALIAÇÃO HEURÍSTICA	59
4.2	TESTES DE USABILIDADE	62
5	CONCLUSÕES E TRABALHO FUTURO	67
5.1	RESUMO	67
5.2	PRINCIPAIS CONTRIBUIÇÕES	68
5.3	TRABALHO FUTURO	68
	BIBLIOGRAFIA	71
APÊNDICE A	CONVERSÃO DE DADOS GEOGRÁFICOS	75
	APÊNDICE A.1 CONVERSÃO DAS SHAPEFILES PARA TOPOJSON	75
	APÊNDICE A.2 CONVERSÃO PARA GEOJSON	77
APÊNDICE B	MIGRAÇÃO DE DADOS	79
APÊNDICE C	TESTES DE USABILIDADE: DOCUMENTOS	81
	APÊNDICE C.1 GUIA DE TAREFAS DO UTILIZADOR	81
	APÊNDICE C.2 FICHA DO OBSERVADOR	87
	ANEXOS	89

Lista de Tabelas

Tabela 2.1 Representação em matriz adjacência de dados OD representando o movimento dos candidatos ao ensino superior do distrito de candidatura para o distrito de colocação.	11
Tabela 2.2 Lista de adjacências em tabela com múltiplos atributos.....	11
Tabela 3.1 Descrição dos atributos do candidato.	31
Tabela 3.2 Descrição dos atributos do GAES.	31
Tabela 3.3 Descrição dos atributos do estabelecimento de colocação.	32
Tabela 3.4 Descrição dos atributos do distrito.	33
Tabela 3.5 Descrição do conteúdo dos ficheiros otimizados.....	36
Tabela 3.6 Indicadores utilizados.	36
Tabela 3.7 Exemplo simplificado dos fluxos entre distritos.	51

Lista de Figuras

Figura 2.1 Lista numérica sem recurso a qualquer atributo visual (Freeman, 2015).	6
Figura 2.2 Lista numérica da Figura 2.1 com destaque do número 3 através da cor (Freeman, 2015).....	6
Figura 2.3 Teste de precisão através de posição-ângulo (Cleveland e McGill, 1984).	6
Figura 2.4 Mapa de John Snow: Local de morte por cólera (Chen, 2010).....	7
Figura 2.5 Papel da comunicação visual (Freeman, 2015).....	8
Figura 2.6 Pipeline de Visualização de Informação, adaptado de (Liu <i>et al.</i> , 2014).....	9
Figura 2.7 (a) Grafo orientado; (b) Lista de adjacências; (c) Matriz de adjacências (Cormen <i>et al.</i> , 2001).....	10
Figura 2.8 Representação visual da matriz de adjacência das tabelas anteriores.	12
Figura 2.9 Excerto da visualização de Harness - Fluxo de passageiros nos caminhos de ferro em 1837 (Robinson, 1955).	13
Figura 2.10 Mapa figurativo de padrões de migrações internacionais em 1858 (Minard, 1862)..	14
Figura 2.11 - Mapa com o movimento das notas de um dólar nos EUA (1965-1970) (Tobler, 1984).....	15
Figura 2.12 Balanço das migrações dominantes nos EUA de 1995-2000.....	15
Figura 2.13 - Visualização de migração do estado do Colorado para outros estados dos EUA (1995-2000) (Doantam Phan <i>et al.</i> , 2005).	16

Figura 2.14 Grafo desenvolvido para representar a migração dos candidatos do Ensino Superior, no ano de 2012.....	17
Figura 2.15 - Visualização em matricial das migrações entre os concelhos regionais da Nova Zelândia (Bryant, 2011).....	18
Figura 2.16 Diagrama de Sankey dos candidatos do ensino superior em 2014.	19
Figura 2.17 Secção das das origens dos candidatos em promenor.	19
Figura 2.18 Diagrama de Sankey em arco. (Nagel <i>et al.</i> , 2012).....	20
Figura 2.19 –Gráfico circular do fluxo de migratório entre as várias regiões do mundo. (Abel e Sander, 2014).....	21
Figura 2.20 - Exemplo de mapeamento do domínio numérico para escala de cores (Meeks, 2015).	22
Figura 2.21 - Diagrama de projeções desenvolvido por (Jenny, 2012).	23
Figura 2.22 Google Chrome developer tools. No topo o DOM com os elementos SVG e em baixo o resultado da seleção do D3.	26
Figura 3.1 Excerto dos dados associados aos candidatos (primeiros 9 do ano 2012).	30
Figura 3.2 Excerto da tabela GAES.....	31
Figura 3.3 Excerto dos estabelecimento de colocação.	32
Figura 3.4 Excerto da informação contida na tabela distritos.	33
Figura 3.5 - Diagrama ER do modelo de dados.....	34
Figura 3.6 Exemplo simplificado da migração de candidatos ao Ensino Superior.	35
Figura 3.7 Página inicial do website.....	37
Figura 3.8 Painel de filtros e controlos da visualização de matriz de adjacências.	38
Figura 3.9 Painel de detalhe dos dados do distrito de Aveiro.	38
Figura 3.10 Painel de detalhe dos candidatos de Aveiro colocados em Beja.....	39
Figura 3.11 Lista de estabelecimentos com candidatos provenientes do distrito do Porto colocados em Aveiro.	40
Figura 3.12 Visualização da Matriz de Adjacências.	40
Figura 3.13 Legenda dos distritos de destino (tamanho do texto proporcional à população).	41
Figura 3.14 Escala de cores e legenda relativa ao número de candidatos.	41
Figura 3.15 Nove tonalidades de azul.....	42
Figura 3.16 Os nove tons com os valores dos quantis.....	42
Figura 3.17 Escala do saldo líquido dos fluxos.	42

Figura 3.18 Transição entre os dois modos.	43
Figura 3.19 Matriz no estado de “saldo líquido dos fluxos”.	43
Figura 3.20 Diagrama cordal.	44
Figura 3.21 Detalhe dos fluxos do distrito de Santarém.	45
Figura 3.22 Visualização dos candidatos com valores dos filtros por omissão.	46
Figura 3.23 Visualização dos candidatos do sexo masculino de 2014 do Ensino Superior Público Politécnico.	46
Figura 3.24 Representação de Georreferenciação de fluxos.	47
Figura 3.25 Representação dos 20 distritos com 20 cores.	48
Figura 3.26 Mapeamento da população de cada distrito usando cor.	48
Figura 3.27 - Projeção Albers Portugal criada em D3.js para representar Portugal e regiões autónomas.	49
Figura 3.28 Indicadores associados ao tamanho do círculo de cada distrito	49
Figura 3.29 Mapa de fluxo após ação do utilizador.	50
Figura 3.30 Detalhe dos arcos e círculos	51
Figura 3.31 Interpolação de uma curva de Bézier de P0 para P2.	52
Figura 3.32 Detalhe das curvas de Bézier representando diferentes arcos.	52
Figura 3.33 – Representação do fluxo para o mesmo distrito.	53
Figura 3.34 Ilustração das várias combinações dos parâmetros para o arco elíptico.	53
Figura 3.35 Valor do saldo do fluxo entre dois distritos.	54
Figura 3.36 Número total de candidatos.	54
Figura 3.37 Arco selecionado sobreposto a todos os outros.	55
Figura 3.38 – Ilustração dos passos para a geração de TopoJSON simplificado (3) a partir de um Shapefile (1).	56
Figura 4.1 Matriz antes da avaliação Heurística.	60
Figura 4.2 Alteração das escalas de cores depois da avaliação heurística.	60
Figura 4.3 Filtro do género antes da avaliação.	61
Figura 4.4 Filtro do género depois do aperfeiçoamento.	61
Figura 4.5 Painel de ajuda contextual.	61
Figura 4.6 Resultado das respostas do Grupo 1	63
Figura 4.7 Tempos médios de conclusão das tarefas do Grupo 1.	64

Figura 4.8 Nível de facilidade atribuído pelo utilizador e a visualização mais frequente (Grupo 2).	64
Figura A.0.1 Representação do conteúdo presente na <i>shapefile</i>	75
Figura A.0.2 – Descrição dos metadados associados a polígonos da <i>shapefile</i>	76

1 Introdução

1.1 Motivação

Hoje em dia a maior parte dos acontecimentos são armazenados digitalmente. Estima-se que só em 2010 tenham sido gerados 1,200 *exabytes* de dados e estima-se que esta quantidade continue a aumentar a uma taxa de 59% ao ano (Heer, Bostock e Ogievetsky, 2010).

Este enorme volume de dados contém informações valiosas que nos permitem gerir negócios, empresas e organizações, melhorando a sociedade e o nosso dia-a-dia. Mas para tomarmos as decisões mais adequadas, temos que encontrar forma de explorar, relacionar e comunicar os dados de modo eficiente.

O objectivo da Visualização de Informação é ajudar as pessoas na compreensão desses dados, aproveitando a capacidade do sistema visual humano para detectar padrões, identificar tendências e casos anómalos (Munzner, 2014).

Desde há muito tempo que os computadores revolucionaram a nossa forma de trabalhar e o uso de aplicações interativas revolucionou a forma como interagimos com as aplicações, existindo um interesse crescente na análise visual dos dados de migrações, tendo sido desenvolvida, nos últimos anos, uma variedade de técnicas, métodos e ferramentas para representar e analisar este tipo de informação (Andrienko e Andrienko, 2012).

“O concurso nacional de acesso ao ensino superior é um dos maiores concursos organizados pela Administração Pública, com relevo e destaque na sociedade civil.”

(DGES, 2011)

Este trabalho inserido na área de Visualização de Informação, pretende propor formas de visualizar as migrações dos candidatos colocados na primeira fase do concurso nacional de acesso ao Ensino Superior público de 2012 a 2014. O desafio da utilização de um conjunto de dados real com tamanho significativo e a curiosidade sobre o seu conteúdo pelo facto de também ter sido candidato ao Ensino Superior, proporcionaram a motivação necessária para o desenvolvimento deste trabalho.

1.2 Objetivos

O objectivo principal desta dissertação é desenvolver e avaliar representações interativas para visualizar dados de movimentos migratórios, possibilitando uma melhor compreensão sobre o tema através de um caso de estudo: resultados do concurso nacional de acesso ao Ensino Superior.

A simples representação tabular do número de residentes deslocados para uma dada região ou país, consoante o seu país ou região de origem, não permite perceber, de modo simples, a intensidade de fluxos migratórios nem detectar padrões globais de comportamento.

Esta dissertação deverá também ilustrar o processo completo de criação de representações visuais, desde a obtenção e processamento dos dados, passando pelo desenvolvimento até à validação e avaliação das visualizações desenvolvidas.

Estas visualizações devem ser integradas numa plataforma Web de fácil utilização e consulta, contemplando funcionalidades de filtragem, cálculo de indicadores, agregação de informação complementar e animações intuitivas.

Além disto, esta dissertação pretende demonstrar que a Web e as ferramentas que a compõem sofreram um grande avanço, permitindo agora desenvolver representações visuais complexas e de grande interatividade, melhorando a visualização, percepção e utilização dos dados originais.

1.3 Visão geral

No Capítulo 2 é apresentada uma introdução à Visualização de Dados e à Visualização de Informação, descrevendo as suas principais diferenças e analisando o seu processo de

desenvolvimento. É também feita uma síntese das representações existentes para a visualização de fluxos migratórios.

O Capítulo 3 apresenta o conjunto de dados utilizados como caso de estudo, os resultados da primeira fase do concurso de acesso ao Ensino Superior de 2012 a 2014, descreve o conjunto das representações desenvolvidas, fundamentando as escolhas técnicas efectuadas e as principais contribuições produzidas

No Capítulo 4 é descrito todo o trabalho de validação e avaliação das visualizações e métodos de interação desenvolvidos, através de avaliações heurísticas e testes de usabilidade.

Por fim, no Capítulo 5, são apresentados os resultados e conclusões, assim como sugestões para trabalho futuro.

2 Estado da Arte

Este capítulo contém uma introdução à Visualização de Informação e Visualização de Dados e as principais diferenças entre ambas. Aqui é feita também uma introdução aos fluxos migratórios e conceitos inerentes aos estudo dos dados origem-destino. É ainda realizada uma síntese cronológica das representações mais relevantes de fluxos migratórios, desde ilustrações estáticas até visualizações interativas e animadas.

Por fim, é apresentado um resumo das principais ferramentas de visualização de informação utilizadas hoje em dia, fazendo ainda uma avaliação sobre a que se mostrou mais promissora para a concretização deste trabalho.

2.1 Visualização de Informação

“Visualization’s primary goal is to make it easier for people to understand and use vast amounts of data”.

(Wiss, Carr e Jonsson, 1998)

A Visualização de Informação é uma área de estudo relativamente nova, que está a ganhar cada vez mais atenção em diversos sectores, desde o ambiente académico até ao industrial. Explora o uso de vários métodos visuais, desde ilustrações até representações gráficas interativas produzidas por computador para representar dados abstratos e estimular a cognição (Hong, 2014), amplificar as nossas capacidades (Munzner, 2014), tal como na Figura 2.1 onde identificar todos os números três, é uma tarefa difícil e no entanto, na Figura 2.2 a utilização do atributo cor para destacar os número três torna-se simples para o cérebro identificar sem

qualquer esforço a quantidade de três existentes na imagem. No entanto a nossa capacidade de interpretação não é apenas condicionada pelos atributos visuais, mas também pela nossa capacidade de precisão que também deve ser maximizada, tal como é demonstrado por Cleveland e McGill através da Figura 2.3 onde os utilizadores demonstraram mais facilidade na compreensão dos valores representados no gráfico de barras utilizando a altura, do que no gráfico circular através do ângulo. (Cleveland e McGill, 1984).

```

28049385628406947862485
83922089486208947690187
85098834260928468724859
82382409852468749875220
89485202984850924853290
88452029884529028843528
92842589987458784958784
98597076764674153698742

```

Figura 2.1 Lista numérica sem recurso a qualquer atributo visual (Freeman, 2015).

```

28049385628406947862485
83922089486208947690187
85098834260928468724859
82382409852468749875220
89485202984850924853290
88452029884529028843528
92842589987458784958784
98597076764674153698742

```

Figura 2.2 Lista numérica da Figura 2.1 com destaque do número 3 através da cor (Freeman, 2015).

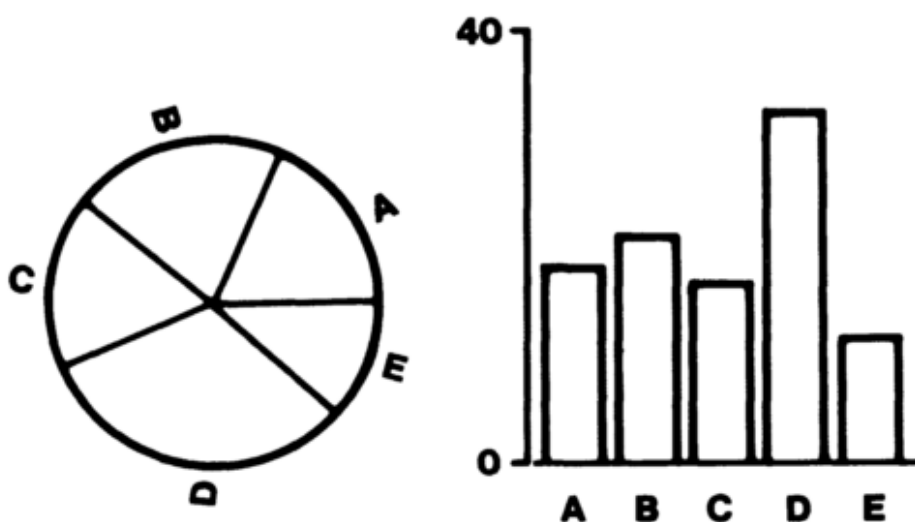


Figura 2.3 Teste de precisão através de posição-ângulo (Cleveland e McGill, 1984).

Qualquer visualização deve por isso garantir, que maximiza a rapidez e a capacidade de decodificar de forma precisa os dados representados. Através da nossa compreensão do

funcionamento da percepção humana, devemos adaptar a informação visual de forma a estimular e incentivar a sua interpretação (Ware, 2004).

2.1.1 Visualização de Informação *versus* Visualização de Dados

Tendo em conta as capacidades do sistema visual humano, torna-se necessário estudar a melhor forma de exibir visualmente dados de diferentes tipos, atendendo à informação contida nos dados e ao tipo de questões que se pretende analisar. A visualização pode categorizar-se em dois tipos Visualização de Informação ou Visualização de Dados. (Chen, 2010) ajuda-nos a compreender a diferença entre estas duas abordagens.

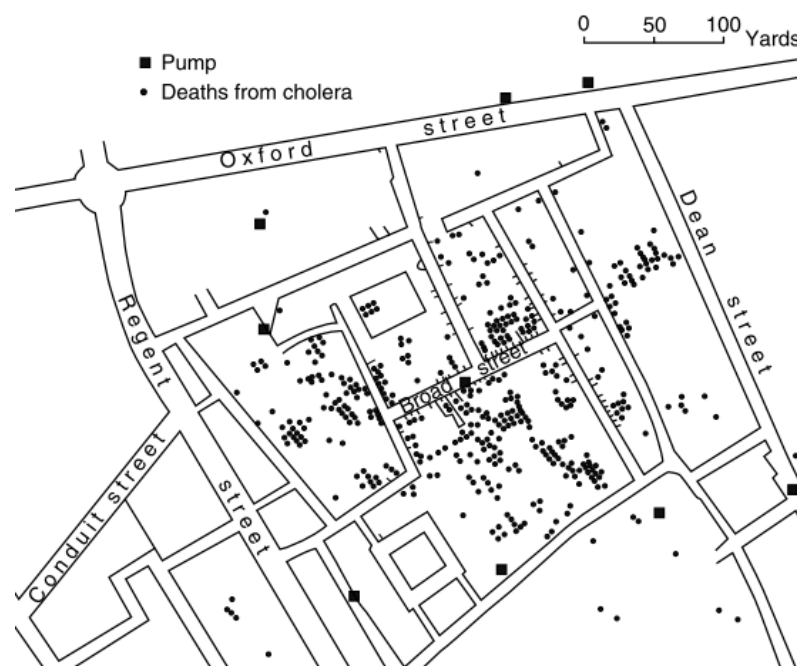


Figura 2.4 Mapa de John Snow: Local de morte por cólera (Chen, 2010).

Através do mapa (Figura 2.4) de criado pelo Dr. John Snow, Chen (2010) explica-nos que foi graças à informação visualizada que se conseguiu chegar à conclusão que a origem das mortes por cólera estava associada a uma fonte de água. Foi a combinação da informação associada ao número de mortos, local de residência e localização das fontes de água da cidade de Londres que o levou à descoberta da origem do problema.

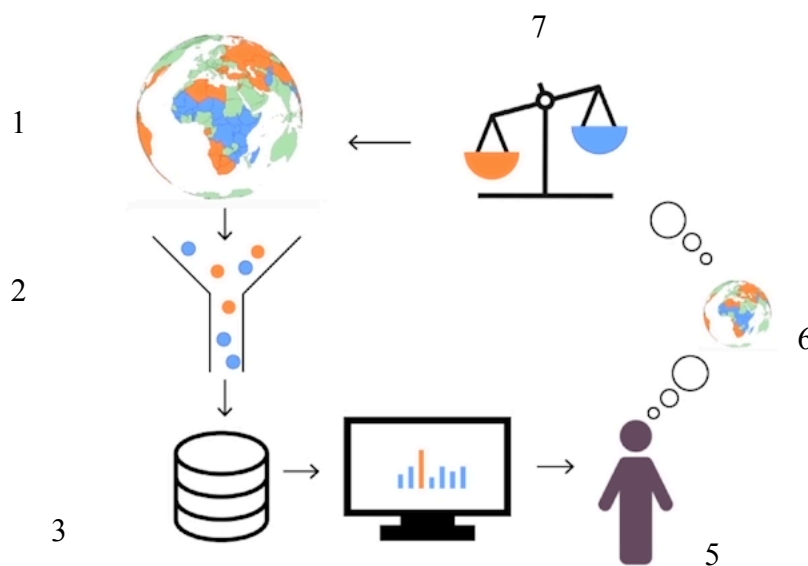


Figura 2.5 Papel da comunicação visual (Freeman, 2015).

(Chen, 2010) explica que a diferença entre visualização de dados e visualização de informação pode ser definida como a presença ou não de dados quantitativos e quão fácil é possível extrair ou transformar os dados em modelos quantitativos.

A visualização de informação não é apenas uma forma de visualizar dados, é sim, um processo completo, tal como representado na Figura 2.5, onde desde a obtenção dos dados (1), processamento, filtragem, transformação e agregação dos dados (2), armazenamento (3), representações visuais (4), que são analisadas por utilizadores (5). Estes utilizadores, por sua vez obtêm o seu modelo mental (6) dos dados através da visualização, que os ajuda a formar conceitos (7) sobre informação que até aqui era demasiado abstrata para analisar (Freeman, 2015).

A Figura 2.6 apresenta os cinco passos principais que tradicionalmente compõem o processo de visualização de informação.

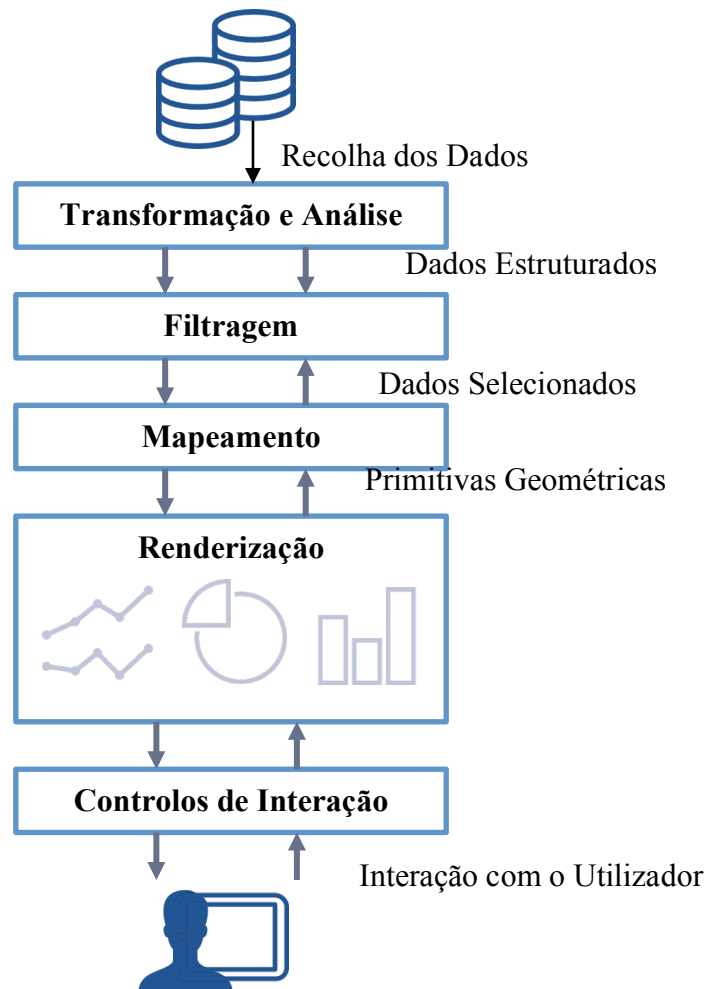


Figura 2.6 Pipeline de Visualização de Informação, adaptado de (Liu *et al.*, 2014).

Na fase de **Transformação e Análise dos Dados**, os dados obtidos são processados e transformados em dados estruturados para facilitar o posterior mapeamento em primitivas gráficas e a sua manipulação.

A **Filtragem** permite reduzir o domínio dos dados, caso exista a necessidade de seleccionar um subconjunto dos dados iniciais para, por exemplo, evitar problemas de desempenho.

Na fase de **Mapeamento** os dados triados na fase de filtragem são representados usando primitivas visuais (e.g., pontos, linhas) e os seus atributos mapeados em características dessas primitivas (e.g., cor, posição, tamanho).

A fase de **Renderização** é responsável por representar de modo apropriado as primitivas geométricas e construir as representações visuais dos dados, i.e., as correspondentes visualizações.

Os utilizadores podem interagir habitualmente com as representações geradas através de vários **Controlos de Interação**, que permitem a exploração e a análise dos dados através de várias perspectivas melhorando a sua compreensão. Para mais informação consultar as *surveys* mais recentes (Liu *et al.*, 2014) e (Chen, 2010).

O trabalho descrito nesta dissertação irá focar-se sobretudo nas duas últimas componentes (Renderização e Controlos de Interação). No entanto, uma vez que não existe qualquer trabalho prévio de recolha, filtragem e transformação dos dados, é fundamental que este trabalho incida também sob todas as fases descritas anteriormente, completando o ciclo de desenvolvimento de Visualização de Informação.

2.2 Introdução aos fluxos migratórios e dados Origem-Destino

Ao contrário da fertilidade e da mortalidade, a migração não ocorre devido a um processo biológico. Pode ser motivada por alterações tecnológicas, eventos, etc. Por este motivo, a compreensão e análise destes movimentos são cada vez mais importantes para a nossa sociedade.

Os dados de fluxos migratórios são um subconjunto dos dados origem-destino, uma vez os fluxos migratórios estão associados à transição específica de seres vivos, de uma origem para um destino. Quando estes movimentos ocorrem dentro do próprio país, são denominados por migrações internas.

Dados do tipo origem-destino (OD) representam relações espaciais, ou seja, fluxos de entidades entre pares de localizações geográficas. Nos dados OD são especificados as origens, os destinos e as magnitudes dos fluxos.

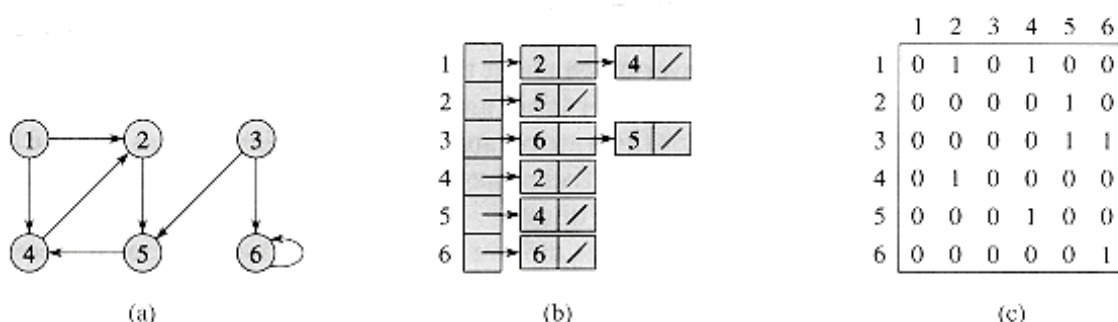


Figura 2.7 (a) Grafo orientado; (b) Lista de adjacências; (c) Matriz de adjacências (Cormen *et al.*, 2001).

Os dados OD são frequentemente representados e armazenados como uma matriz OD. Na matriz OD as linhas e as colunas correspondem às origens e aos destinos, respectivamente (na Tabela 2.1 que são especificadas como distritos) e os valores nas células da matriz representam as magnitudes dos fluxos entre as origens e respectivos destinos. Esta é também conhecida por representação em matriz de adjacência do correspondente grafo orientado (Cormen *et al.*, 2001).

Origem \ Destino	Aveiro	Beja	Braga	Bragança	...
Aveiro	890	0	32	4	...
Beja	5	71	1	0	...
Braga	102	0	2318	32	...
Bragança	7	0	30	72	...
Castelo Branco	12	1	7	1	...
Coimbra	59	0	3	3	...
Évora	1	11	6	0	...
Faro	12	18	14	2	...
Guarda	32	0	6	3	...
Leiria	82	0	18	3	...
...

Tabela 2.1 Representação em matriz adjacência de dados OD representando o movimento dos candidatos ao ensino superior do distrito de candidatura para o distrito de colocação.

Em muitas situações, não existe apenas um único atributo associado ao fluxo, ou seja, os arcos poderão representar múltiplos atributos. Nestes casos a maneira mais fácil para armazenar estes dados é sob a forma de uma lista de adjacências.

Origem	Destino	Atributo 1	Atributo 2	...
Aveiro	Aveiro	890	2012	..
Aveiro	Beja	0	2012	..
Aveiro	Braga	32	2013	...
Aveiro	Bragança	4	2014	...
...

Tabela 2.2 Lista de adjacências em tabela com múltiplos atributos.

Graficamente, estes dados OD podem ser visualizados em forma de matriz como apresenta a Figura 2.8.

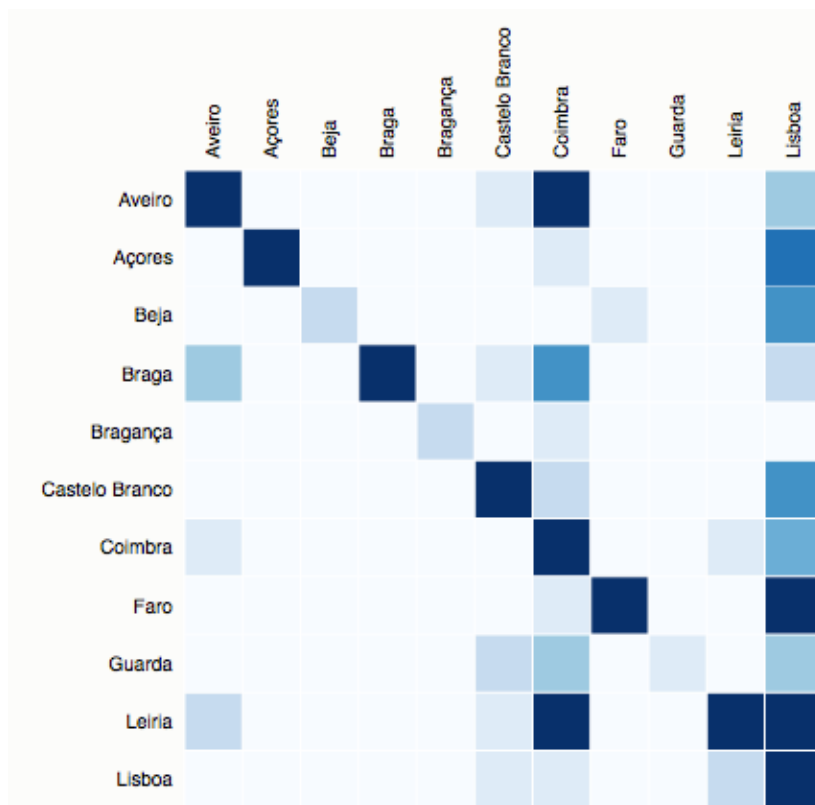


Figura 2.8 Representação visual da matriz de adjacência das tabelas anteriores.

Os dados de origem-destino encontram-se estruturados ou são facilmente representados por uma matriz. Dada esta interpretação, a transformação da matriz de adjacências numa representação visual acaba por ser intuitiva. Usar uma escala de cores em vez de valores permite que os valores associados aos arcos possam ser percebidos mais rapidamente. (Heer *et al.*, 2010)

2.3 Visualização de Fluxos Migratórios

“A grande parte das alterações mundiais estão associadas a migrações geográficas”.

(Tobler, 2005)

2.3.1 Evolução

Como grande parte das visualizações existentes hoje em dia, a visualização de fluxos migratórios não é uma novidade na área de visualização de informação. A primeira visualização de fluxos origem-destino conhecida foi criada por Henry Harness em 1837. A Figura 2.2 mostra o mapa de

fluxos origem-destino elaborado por Harness para os *Railway Commissioners* para representar o número dos passageiros transportados nas diferentes direções pelos caminhos de ferro da Irlanda. (Robinson, 1955) alega que foi a primeira utilização de uma linha sobre um mapa com espessura proporcional ao valor numérico dos dados.

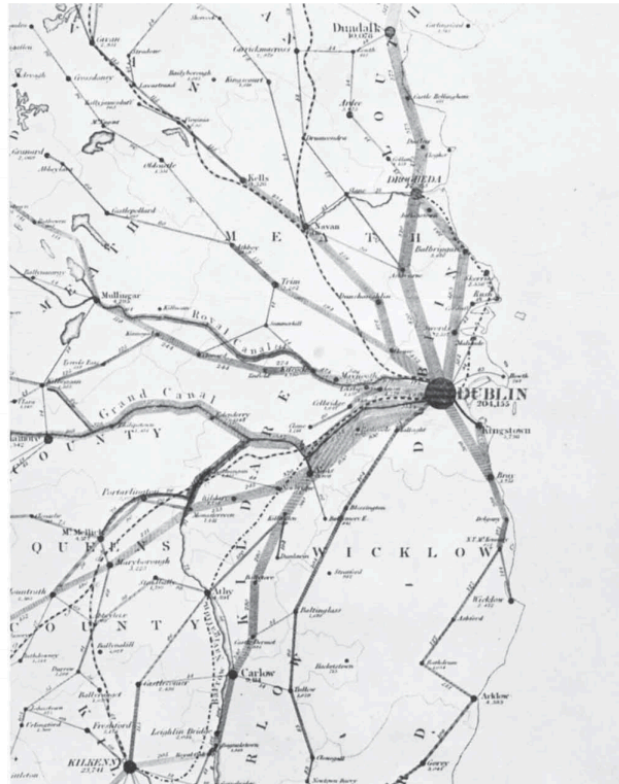


Figura 2.9 Excerto da visualização de Harness - Fluxo de passageiros nos caminhos de ferro em 1837 (Robinson, 1955).

Posteriormente, entre 1845 e 1869 o francês Charles Minard ficaria na história da Visualização de Informação por popularizar a mesma técnica aplicando-a em múltiplas visualizações.

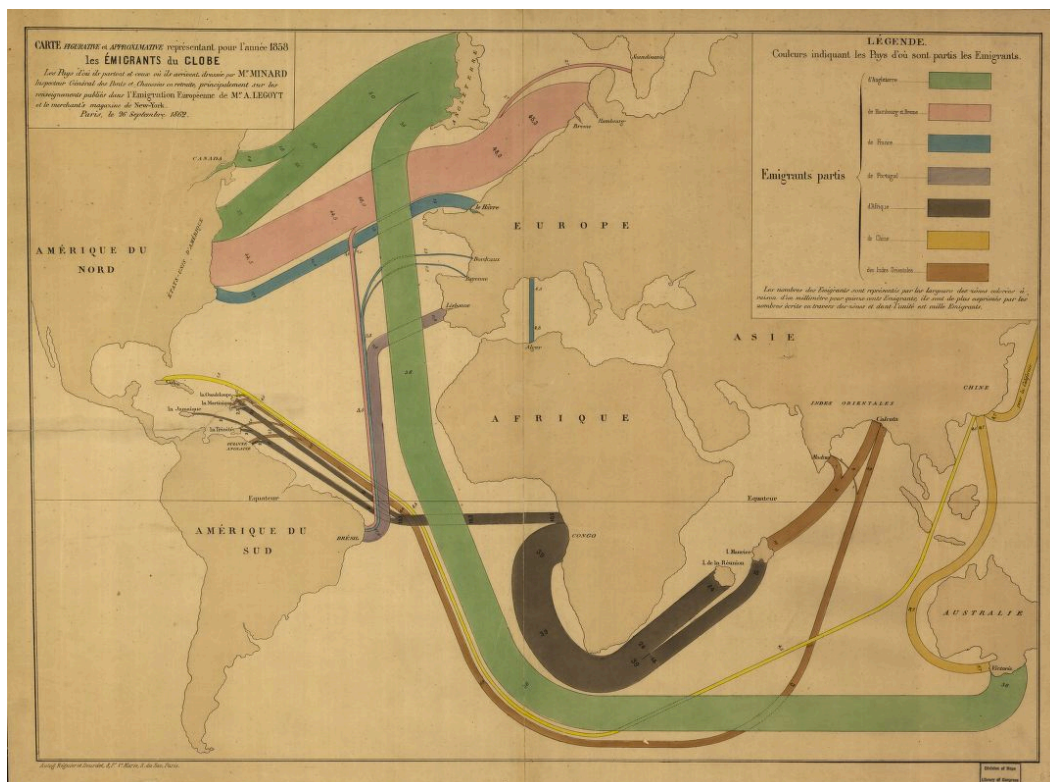


Figura 2.10 Mapa figurativo de padrões de migrações internacionais em 1858 (Minard, 1862).

A Figura 2.10 representa um mapa de 1858 criado por Minard que é ainda hoje uma das suas obras mais emblemáticas na área de Visualização de Informação e provavelmente a primeira a representar migrações a nível global.

Os primeiros fluxos migratórios representados sobre mapas, desenvolvidos com recurso a um computador datam das décadas de 1960 e 1970. Eram principalmente utilizados por engenheiros de transportes, mas mais tarde tiveram aplicações noutros domínios e.g. cuidados de saúde (Francis e Schneider, 1984).

Já nos anos 1980 e 1990 foram desenvolvidas várias ferramentas de software para representação de fluxos migratórios georreferenciados. Uma das ferramentas mais mencionadas nas publicações científicas sobre fluxos migratórios foi o *Flow Mapper*¹ desenvolvido pelo professor e cartógrafo Waldo Tobler (Boyandin *et al.*, 2010). Nas Figura 2.11 e Figura 2.12 podemos observar duas representações geradas através da ferramenta *Flow Mapper* em versões diferentes,

¹ <http://www.csiss.org/clearinghouse/FlowMapper/> (consultado em 9-07-2015)

a versão de 1984 sem qualquer cor ou interação e a versão de 2000 com algoritmos melhorados, cor e interação.

A Figura 2.11 representa os fluxos do movimento das notas de um dólar gerada com a primeira versão do *Flow Mapper*, e a Figura 2.12 o balanço simplificado do total das migrações nos Estados Unidos da América gerado com uma versão mais recente do mesmo software.

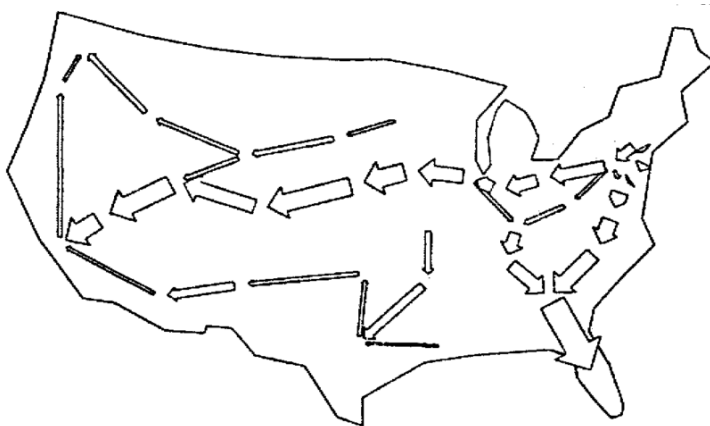


Figura 2.11 - Mapa com o movimento das notas de um dólar nos EUA (1965-1970) (Tobler, 1984).

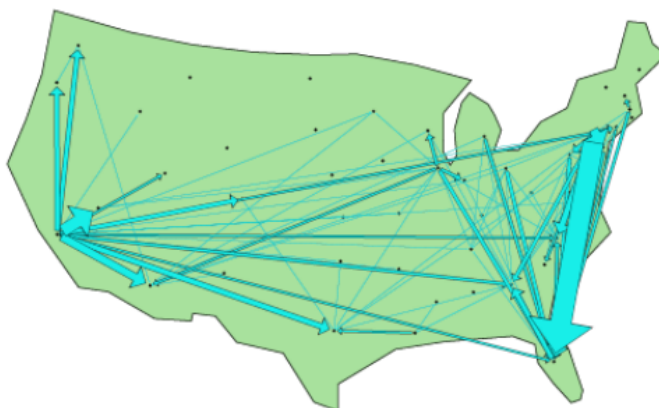


Figura 2.12 Balanço das migrações dominantes nos EUA de 1995-2000².

O artigo de (Tobler, 1984) sobre visualizações de fluxos em mapas, apresenta-nos um conjunto de padrões de movimentos geográficos bastante eficazes como ferramentas ilustrativas e de investigação já criados com o auxílio de computadores. Apesar do salto tecnológico, o conceito principal destas visualizações não se alterou muito desde a época de Minard no século XIX, onde

² <http://www.csiss.org/clearinghouse/FlowMapper/FlowTutorial.pdf> (consultado em 9-07-2015)

as regiões entre as quais ocorre uma migração, ou outro movimento, são ligadas por uma "faixa", cuja largura representa a quantidade transferida.

2.4 Métodos de visualização de fluxos migratórios

A evolução das tecnologias permite cada vez mais captar e armazenar dados com informação geográfica ou origem-destino associados, esta tendência irá aumentar nos próximos anos, quer por razões tecnológicas como razões sociais (Keim, Panse e Sips, 2004). Esta tendência tem originado um interesse por parte da comunidade científica e empresarial, como poderemos ver pelos vários exemplos de visualizações mais recentes e populares.

2.4.1 Fluxos georreferenciados

Os métodos atuais para representação deste tipo de fluxos baseiam-se quer no trabalho precursor de C. Minard (Minard, 1862) quer em trabalhos mais recentes como o de W. Tobler (Tobler, 1984, 2005), e reconhecem a mais valia da representação de fluxos em mapas de forma a combinar a informação dos dados com a localização geográfica.

A Figura 2.13 apresenta um exemplo de utilização de um conjunto de técnicas, como filtros, agregação dos fluxos e interação com o utilizador com a finalidade de melhorar e simplificar a análise e a compreensão de fluxos georreferenciados.

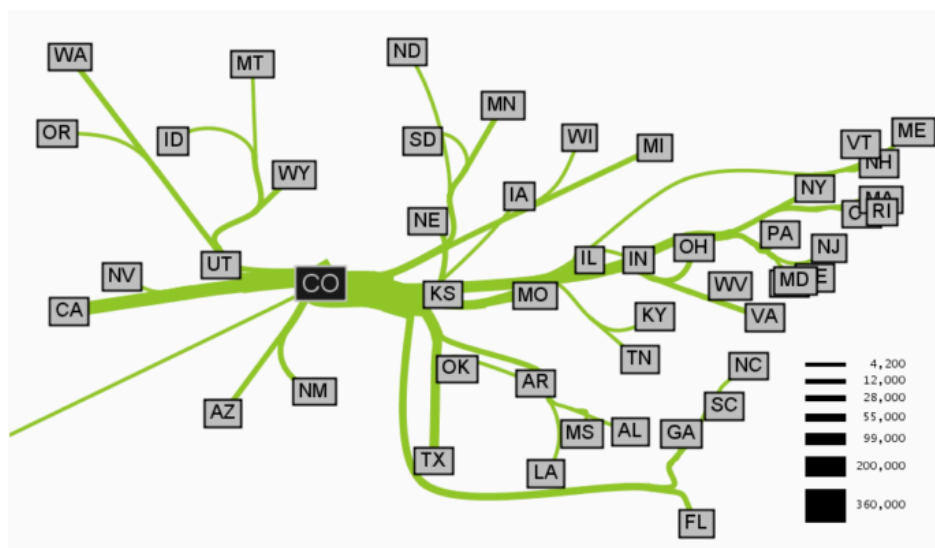


Figura 2.13 - Visualização de migração do estado do Colorado para outros estados dos EUA (1995-2000) (Doantam Phan *et al.*, 2005).

2.4.2 Grafo

Os grafos oferecem uma representação intuitiva e simples para visualizar os fluxos migratórios. As origens e destinos são representados por nós e os fluxos entre a origem e destino são representados através de arcos entre ambos como explica (Heer *et al.*, 2010).

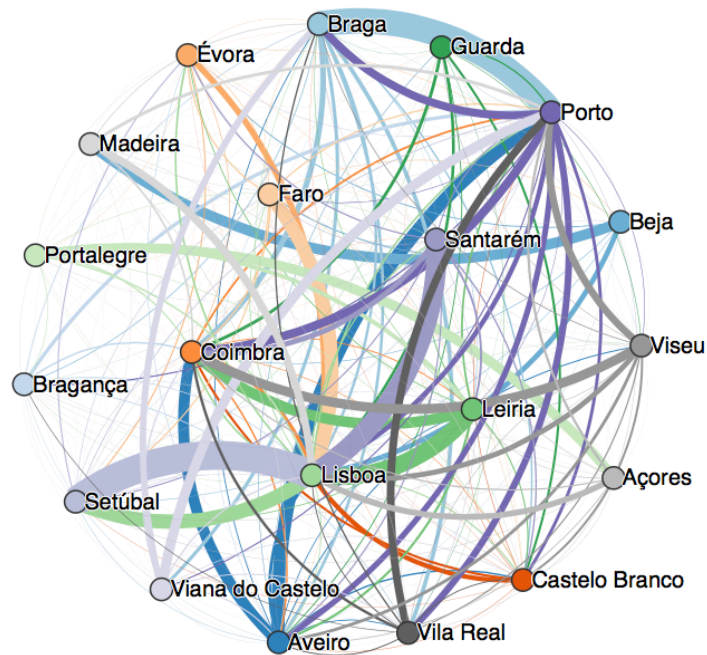


Figura 2.14 Grafo desenvolvido para representar a migração dos candidatos do Ensino Superior, no ano de 2012.

O mais comum, tal como acontece na Figura 2.14 é que estes grafos sejam orientados e pesados, estes atributos normalmente mapeados visualmente através da cor e espessura do arco respectivamente.

2.4.3 Representação Matricial

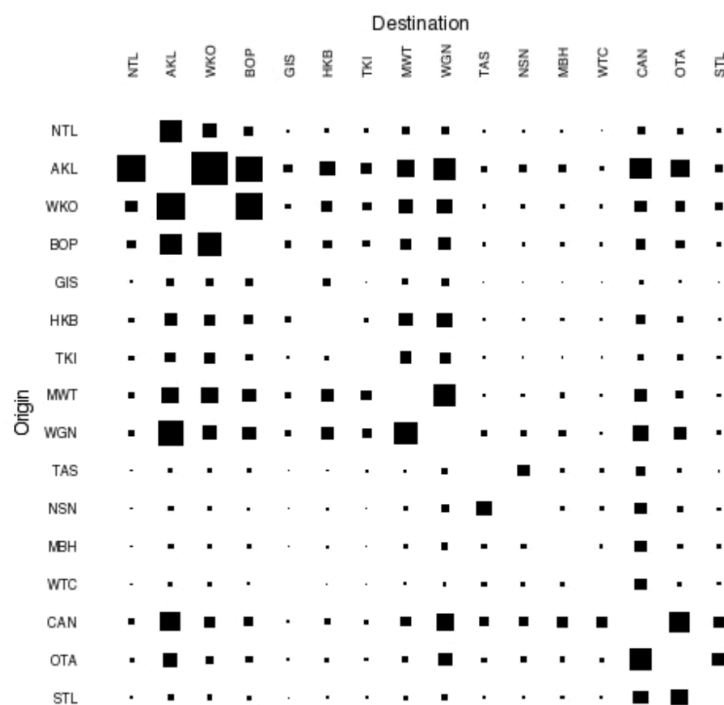


Figura 2.15 - Visualização em matricial das migrações entre os concelhos regionais da Nova Zelândia (Bryant, 2011).

John Bryant reconhece a dificuldade de representar visualmente grandes quantidades de dados de fluxos migratórios, nomeadamente quando cada origem está ligada a todos os destinos (grafo completo). Tendo em conta essa dificuldade propõe a visualização da Figura 2.15, onde os fluxos estão representados numa matriz e a magnitude do fluxo entre a origem e destino é proporcional ao tamanho do quadrado do elemento correspondente da matriz.

Através desta representação (Bryant, 2011) mostra-nos que, não só é viável representar grandes quantidades de fluxos, como também os utilizadores estão mais familiarizados com este modelo visual, devido à sua semelhança com o modelo tabular.

2.4.4 Diagrama de Sankey

Já no decorrer do desenvolvimento desta dissertação, a Direção-Geral do Ensino Superior (DGES) inaugurou um site que apresenta estatísticas sobre os candidatos ao ensino superior em 2014 que também serviram de caso de estudo para este trabalho. Na plataforma é possível consultar as “Preferências por distrito ou região autónoma de candidatura e de permanência ou

mobilidade” através da Visualização mostrada na Figura 2.16 e em pormenor na Figura 2.17. Esta visualização é uma implementação direta do diagrama de *Sankey* clássico, onde do lado esquerdo estão os distritos de origem e do lado direito os distritos de destino. A espessura dos arcos que unem as duas extremidades representa a quantidade de candidatos que transita de um distrito para outro. Este site e esta visualização em particular demonstram o interesse crescente pela visualização e análise de fluxos migratórios uma vez que se começa a migrar de representações tabulares para visualizações interativas e animadas, optimizadas para a Web como explica (Ho *et al.*, 2011).

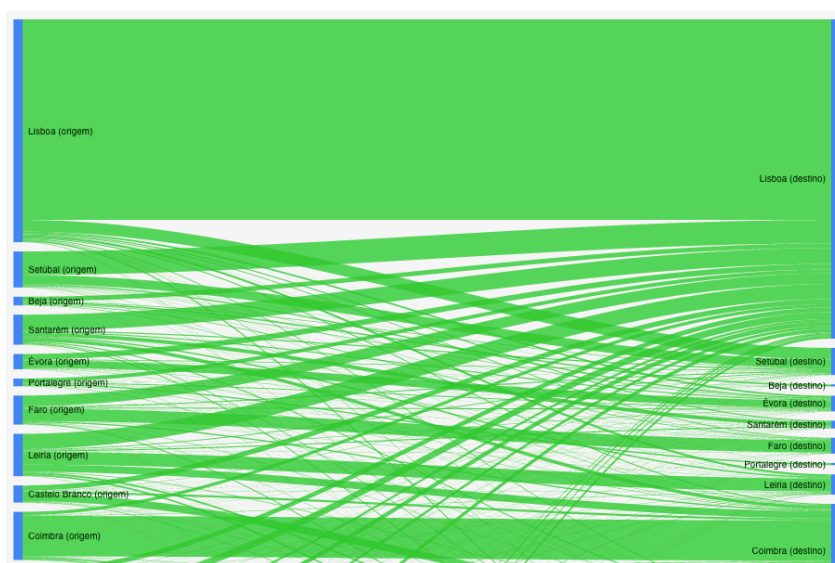


Figura 2.16 Diagrama de Sankey dos candidatos do ensino superior em 2014³.

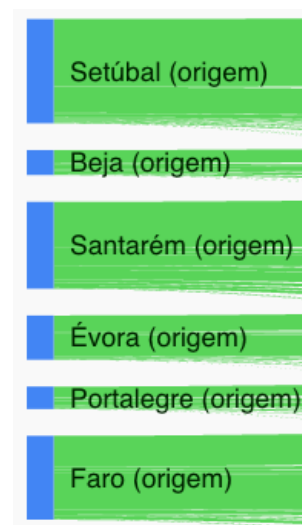


Figura 2.17 Secção das das origens dos candidatos em pormenor.

No entanto, tal como noutras abordagens através de diagramas de Sankey verticais como é o caso do site PeopleMov - *migration flows across the world*⁴ de Carlo Zapponi, as representações verticais apresentam problemas de usabilidade, uma vez que o percurso do fluxo pode não ser completamente visível no ecrã devido à distância entre as origens e destinos.

Para evitar este problema, (Nagel *et al.*, 2012) sugere a utilização de diagramas de Sankey em arco como ilustrado na Figura 2.18. Este método consegue uma representação mais compacta e

³ http://www.dges.mec.pt/estatisticasacesso/cand_13.php (consultado em 9-07-2015)

⁴ <http://peoplemov.in/> (consultado em 9-07-2015)

clara, permitindo aos utilizadores compreender e comparar os pesos dos arcos e ao mesmo tempo observar o trajeto do fluxo.

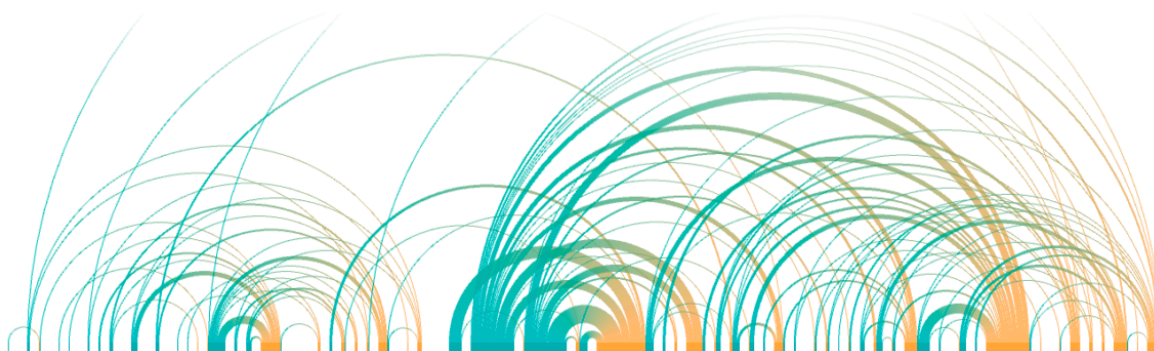


Figura 2.18 Diagrama de Sankey em arco. (Nagel *et al.*, 2012).

2.4.5 Diagrama cordal

Os diagramas cordais não utilizam qualquer representação geográfica e utilizam um *layout* circular para representar os locais. As transições entre origens e destinos são representadas por arcos de espessura variável que ligam os dois nós. Através dos arcos são expostas as magnitudes do fluxo em ambas as direções, uma vez que a espessura do arco pode ser diferente nas duas extremidades.

Nikola Sander e a sua equipa de investigação, é um desses casos mais recentes, onde ao longo dos últimos anos, têm desenvolvido várias visualizações de fluxos migratórios tendo como base esta técnica de representação, ilustrada na Figura 2.19.

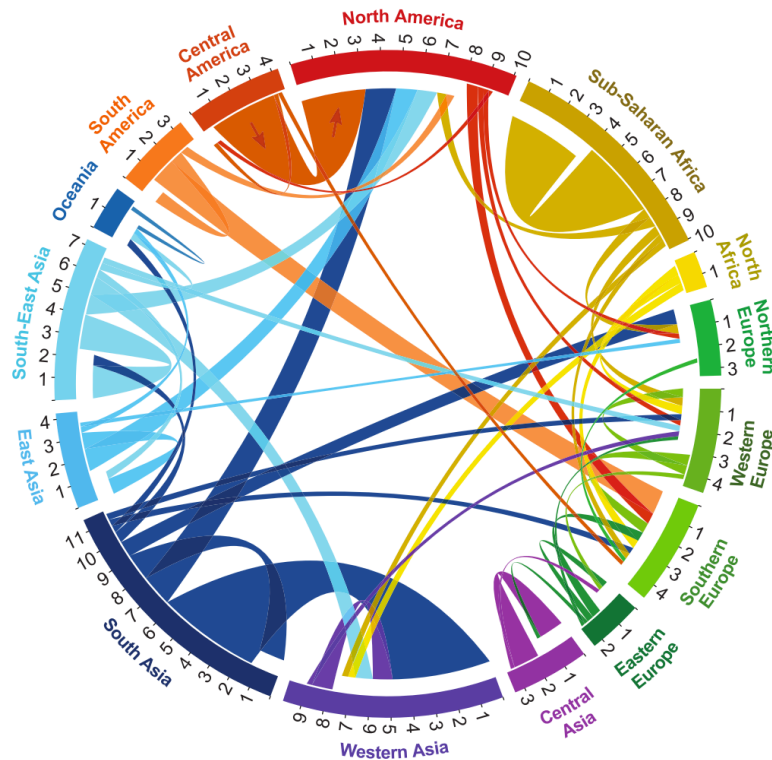


Figura 2.19 –Gráfico circular do fluxo de migratório entre as várias regiões do mundo. (Abel e Sander, 2014).

Esta visualização é particularmente interessante para ter uma visão intuitiva sobre as relações dos diferentes fluxos, das suas quantidades e compreender a sua proporção relativa, uma vez que a espessura dos arcos é proporcional ao valor do fluxo.

Para além das representações anteriormente descritas, existem muitas outras que podem ser consultadas nos mais recentes *surveys* como (Chen, 2010; Heer e Shneiderman, 2012; Liu *et al.*, 2014), no entanto estas foram escolhidas porque melhor se adequam ao caso de estudo desta dissertação.

2.4.6 Escalas

As escalas permitem transformar um valor numérico de um determinado domínio ou intervalo de origem noutra valor numérico de outra gama ou intervalo de destino.

No capítulo sobre escalas, (Murray, 2013) alerta que num determinado conjunto de dados é improvável que os valores correspondam exatamente às medidas em *pixels* disponíveis para os representar. As escalas proporcionam uma forma conveniente de mapear os valores dos dados para novos valores mais convenientes à visualização.

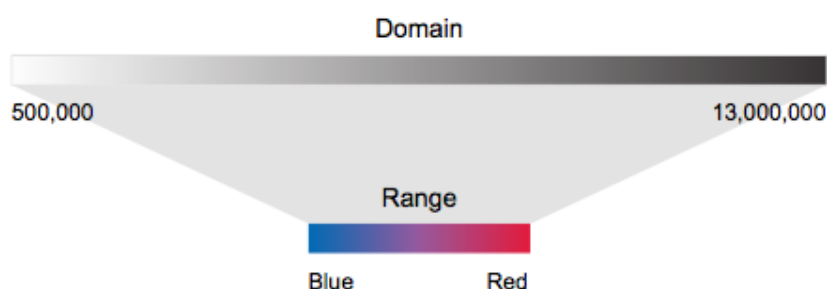


Figura 2.20 - Exemplo de mapeamento do domínio numérico para escala de cores (Meeks, 2015).

Este trabalho, devido à quantidade de dados e heterogeneidade dos valores, faz uso de vários tipos de escalas, tais como, lineares, ordinais e logarítmicas entre outras. A utilização das escalas não está só associada a mapeamentos numéricos, estas podem ser utilizadas para mapear cores e.g. Figura 2.20.

2.4.7 Utilização de cor

A cor tem um papel crucial na visualização de informação, associar cor à informação visível apesar de à primeira vista parecer uma prática bastante simples, torna-se quase sempre uma das tarefas mais complexas para conseguir um bom resultado final e escolher a cor mais correta para o local correto (Tufté, 1990).

A cor quando bem usada melhora e clarifica qualquer visualização. Artigos como os de (Stone, 2006), (Few, 2008) e (Silva, Sousa Santos e Madeira, 2011) auxiliaram na utilização de cor, sensibilizando para a sua utilização de forma adequada, evitando inconsistências perceptuais, definindo conjuntos de cores sequenciais, divergentes e qualitativos e a não utilização de cores indistinguíveis para daltónicos. Mesmo utilizando um vasto conjunto de linhas de orientação para a definição de regras na escolha das cores, existem sempre exceções, quer pelo tipo de público, contexto, cultural ou outro (Silva *et al.*, 2011).

2.4.8 Cartografia

Este trabalho, devido aos fluxos georreferenciáveis, exige um conjunto de elementos cartográficos de Portugal continental e Regiões Autónomas. Infelizmente ainda não existe trabalho desenvolvido para o caso de Portugal continental e R. A. em formato compatível com a Web e com o detalhe necessário para esta dissertação. Para tal, foi realizada uma pesquisa para compreender a maneira mais adequada de produzir essa representação.

A visualização de elementos geográficos está dependente da definição de uma projeção. Apesar de existirem uma variedade de projeções do nosso globo, existem projeções mais indicadas a determinados contextos, posições no globo ou morfologia da área geográfica. A projeção de Albers foi escolhida pelas características referidas por (Sakshuwong e Angeli, 2013). Neste caso foram determinantes a posição de Portugal no globo e o compromisso de preservação das áreas sem distorção como é explicado no artigo de (Jenny, 2012).

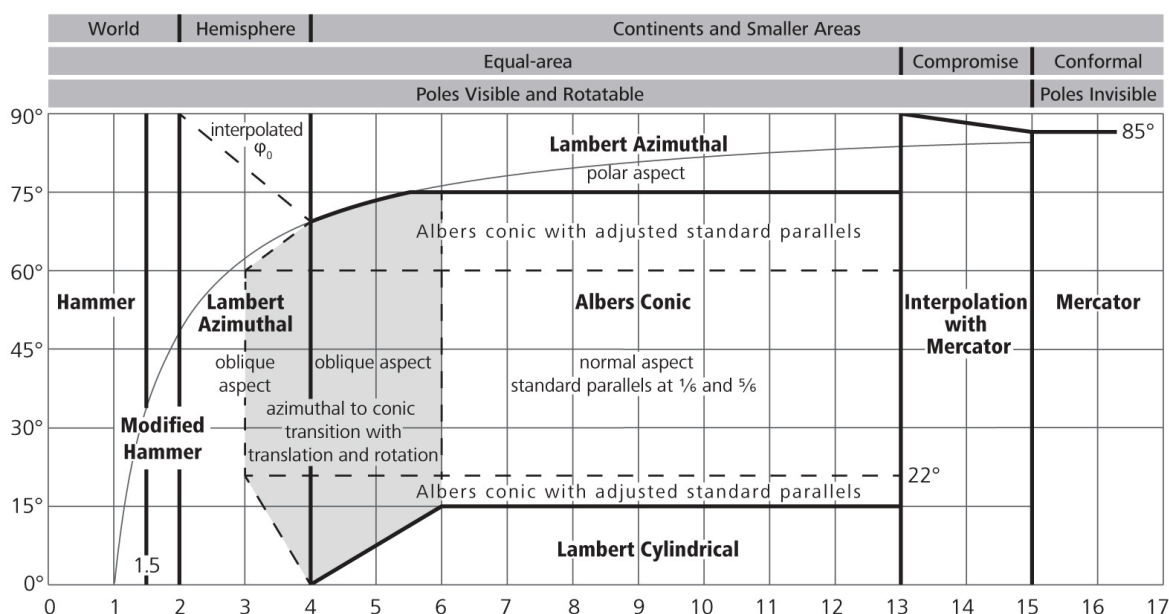


Figura 2.21 - Diagrama de projeções desenvolvido por (Jenny, 2012).

2.5 Ferramentas

Actualmente existe uma variedade enorme de ferramentas disponíveis em diferentes linguagens para criar visualizações.

Entre as possíveis tecnologias e bibliotecas para implementar os vários tipos de visualizações, encontramos ferramentas para desenvolvimento de aplicações de *desktop* e geração de visualizações estáticas, como Circos⁵, R⁶ e Gephi⁷. Por outro lado existe também uma

⁵ <http://circos.ca/> (consultado em 9-07-2015)

⁶ <http://www.r-project.org/> (consultado em 9-07-2015)

⁷ <https://gephi.github.io/> (consultado em 9-07-2015)

diversidade de bibliotecas direcionadas para o desenvolvimento web⁸ e execução no contexto de um browser, das quais se destacam D3.js⁹, HighCharts¹⁰, Raphaël¹¹, entre outras.

Apesar das semelhanças entre as ferramentas em cada um dos grupos anteriormente referidos, a ferramenta escolhida foi a única capaz de satisfazer os critérios expostos na próxima secção, considerados essenciais para este trabalho.

2.5.1 D3.js

A biblioteca D3.js foi criada para colmatar a necessidade permanente de visualizações sofisticadas e ao mesmo tempo compatíveis com a Web. Devido ao seu design robusto, e abstrato, vai muito para além de uma biblioteca de gráfica, antes pelo contrário, a sua implementação é agnóstica de representações visuais pré-definidas (Bostock, Ogievetsky e Heer, 2011).

Especializada e flexível

A D3.js é uma ferramenta para visualização de informação, inclui vários *layouts* gerais para visualizações pré-definidas e contém vários auxiliares de apoio à criação de uma visualização (e.g. mapeamento de escalas, mapeamento de cores, componentes geográficas). No entanto, apesar de ser uma ferramenta de visualização a D3.js é também uma ferramenta especializada em manipulação de dados, facilitando a sua leitura e processamento.

Além disso, é também uma ferramenta profundamente flexível, que permite a criação de visualizações sem qualquer limitação quanto ao tipo de representação gráfica ou estrutura de dados a utilizar.

Interativa

Visualizações interativas e dinâmicas, permitem os utilizadores explorar os dados de forma autónoma. Segundo (Hong, 2014), existe uma relação forte entre o que os utilizadores vêem e como isso influencia o seu processo cognitivo, dando como exemplos os populares clichés de “Uma imagem vale mais que mil palavras” ou “ver para crer”.

⁸ <http://www.sitepoint.com/15-best-javascript-charting-libraries/> (consultado em 9-07-2015)

⁹ <http://d3js.org/> (consultado em 9-07-2015)

¹⁰ <http://www.highcharts.com/> (consultado em 9-07-2015)

¹¹ <http://dmitrybaranovskiy.github.io/raphael/> (consultado em 9-07-2015)

As funcionalidades mais importantes das visualizações interativas são descritas por (Heer e Shneiderman, 2012) como o *zoom*, *pan*, cores, tipo de letra, filtragem e animação.

A utilização da interação e de animações são uma boa forma de captar a atenção dos utilizadores que de outra forma (através dos dados ou visualizações estáticas) não se iriam identificar com o tópico ou dados apresentados (Murray, 2013).

Baseada na Web

As visualizações são feitas para serem visualizadas e a Web é a plataforma mais abrangente hoje em dia, desde computadores de diferentes sistemas operativos, *tablets* e telemóveis todos têm em comum a possibilidade de visualizar o conteúdo existente na Web através do *Browser*.

A D3.js é uma biblioteca escrita em JavaScript e como tal beneficia de todas as características associadas a esta linguagem como explicado por (Heer e Bostock, 2010a) e a sua compatibilidade nativa com a generalidade dos *browsers* Web (Bostock *et al.*, 2011). Esta característica permite que o resultado deste trabalho seja visualizado com uma boa performance, independentemente do sistema operativo, maximizando desta forma o número de utilizadores capazes de visualizar conteúdo produzido através da D3.js.

Comunidade

A comunidade em torno da biblioteca D3.js é uma das mais ativas, com exemplos variados, livros especializados na aplicação da biblioteca. Para este trabalho foram utilizados (Murray, 2013) e (Zhu, 2013) como livros de apoio à biblioteca D3.js.

Orientada ao DOM (*Document Object Model*)

As páginas Web estão estruturadas de acordo com o *Document Object Model*, ou DOM. Este é constituído por uma hierarquia de elementos, designada de nós como apresenta a Figura 2.22.

Existem três categorias de características para cada um dos elementos que determinam o seu comportamento e aparência: **estilos**, **atributos** e **propriedades**. Os estilos podem determinar transparência, cor, tamanho e fronteiras, entre outras. Os atributos referem-se a classes e identificadores (IDs), também podem determinar a aparência ou apenas para referência dos nós. As propriedades são estados dos elementos na página, como por exemplo o estado “Activo” ou “Pressionado”.

A familiarização com o DOM é requisito essencial para desenvolvimento com qualquer tecnologia que ocorra no contexto da web e a D3.js não é exceção.

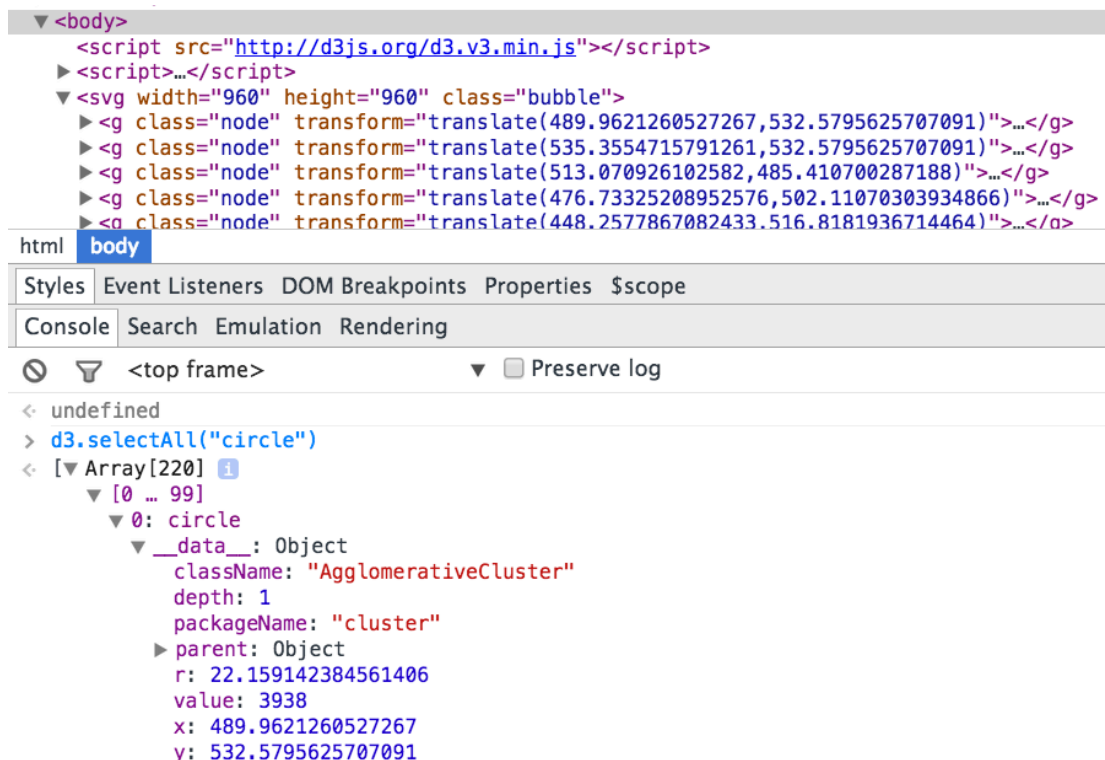


Figura 2.22 Google Chrome developer tools. No topo o DOM com os elementos SVG e em baixo o resultado da seleção do D3.

Apesar da notoriedade desta biblioteca, existe um preconceito infundado quanto à sua dificuldade de utilização. O uso desta ferramenta requer conhecimento prévio sobre um conjunto de tecnologias como HTML, CSS, JavaScript, SVG que caso não exista, irá dificultar a curva de aprendizagem sobre a biblioteca.

2.6 Conclusões

Neste capítulo foi feita uma introdução ao caso de estudo (dados de candidaturas do Ensino Superior português) e aos fluxos migratórios. Foi ainda apresentada a evolução histórica da visualização de fluxos migratórios e também várias representações alternativas para a visualização de fluxos migratórios, no entanto não é possível expor a imensa variedade de representações alternativas existentes para este tipo de problemas.

Para finalizar, foi ainda realizada uma análise às ferramentas existentes, onde a D3.js como muitas bibliotecas de visualização de informação em JavaScript, oferece uma variedade de

funções úteis para abstrair o processo de criação e transformação de elementos de modo a simplificar o trabalho. Além disto, fornece mecanismos para interligar os dados aos elementos da página Web de tal forma que torna o desenho e atualização destes elementos simples e intuitivo para o programador.

No próximo capítulo será apresentado o caso de estudo e o trabalho desenvolvido na sequência desta análise.

3 Visualização das candidaturas ao Ensino Superior Português

Tal como visto no Capítulo 2, criar representações visuais para análise de fluxos é uma tarefa bastante difícil. Este capítulo apresenta-nos o propósito deste trabalho, assim como os dados relativos à 1ª fase do concurso nacional de acesso ao Ensino Superior de 2012 a 2014 utilizados para desenvolver as representações e apresentá-las com recurso a ferramentas atuais em ambiente Web e usando animações, de forma a cativar e auxiliar o utilizador.

3.1 Conjunto de dados

A proposta desta dissertação não contemplava nenhum conjunto de dados em particular e, como tal, foram definidos requisitos que o conjunto de dados deveria satisfazer face aos objectivos da dissertação.

O conjunto de dados deveria possuir as seguintes propriedades:

- Tamanho expressivo, conjuntos demasiados pequenos dificilmente representam um caso real.
- Número significativo de atributos associados aos dados.
- Tema associado interessante, não só para a comunidade académica, mas também para o público em geral.

Após uma análise de várias sugestões de possíveis conjuntos de dados com os orientadores foi decidido, dado o contexto académico, a dimensão dos dados e o número de variáveis envolvidas, utilizar os dados das candidaturas de acesso do Ensino Superior. Na sequência desta decisão, foi

enviado um pedido formal à Direção-Geral do Ensino Superior (DGES)¹², para a autorização e facilitação dos dados necessários ao estudo.

Após a autorização e facultação dos dados, deu-se início ao desenvolvimento técnico descrito nas secções seguintes.

3.1.1 Dados dos Concursos Nacionais de Acesso (1ª fase, 2012 a 2014)

Os dados originais apresentam as colocações dos candidatos da primeira fase do Concurso Nacional de Acesso ao Ensino Superior nos anos de 2012 a 2014. Através destes dados é possível estudar a migração dos candidatos através da relação entre o distrito de candidatura (origem) e o distrito de colocação do candidato (destino).

Os dados encontram-se armazenados num ficheiro Excel que contém 6 folhas de trabalho (2012, 2013, 2014, GAES¹³, Estabelecimentos e Distrito), perfazendo um total de **115.636** candidatos. Nas três primeiras folhas de trabalho, e para o respetivo ano, estão listados os atributos de cada candidato e a instituição em que foi colocado; nas últimas três folhas estão registadas informações complementares relativas às colunas codificadas.

A estrutura dos dados de cada folha de cálculo é descrita pelas figuras e tabelas seguintes.

Candidaturas

A folha contém a informação associada aos vários candidatos para o ano em causa.

	A	B	C	D	E	F	G	H
1	Ano	IDA/luno	FaseCand	GAES	EstabCol	NotaCand	AnoNasc	Sexo
2	2012	1	1	010	0400	132,2	1993	F
3	2012	6	1	010	3062	146	1993	F
4	2012	7	1	010	0300	151,4	1994	F
5	2012	10	1	010	0704	110,5	1992	M
6	2012	13	1	010	3134	171	1994	F
7	2012	16	1	010	0300	162,5	1994	F
8	2012	17	1	010	3014	131	1993	F
9	2012	20	1	010	0300	143,5	1994	F
10	2012	21	1	010	1108	185,7	1994	F

Figura 3.1 Excerto dos dados associados aos candidatos (primeiros 9 do ano 2012).

¹² <http://www.dges.mctes.pt/DGES/pt> (consultado em 9-07-2015)

¹³ GAES - Gabinete de Acesso ao Ensino Superior

Nome	Descrição
Ano	Ano de Candidatura
IDAluno	Identificador único
FaseCand	Fase de candidatura (disponibilizado apenas a primeira fase)
GAES	Código do Gabinete de Acesso Ensino Superior. (i.e. origem do candidato)
EstabalCol	Código do estabelecimento de ensino onde o candidato foi colocado
NotaCand	Nota de candidatura
AnoNasc	Ano de nascimento
Sexo	Género

Tabela 3.1 Descrição dos atributos do candidato.

GAES - Gabinete de Acesso ao Ensino Superior (origem da candidatura)

A folha de cálculo GAES (Gabinete de Acesso ao Ensino Superior) associa ao código de cada GAES a sua localização. Esta é usada para identificar o distrito ou região autónoma de origem dos candidatos.

	A	B
1	GAES	Descr
2	010	Aveiro - Universidade de Aveiro
3	020	Beja - IP de Beja
4	030	Braga - Universidade do Minho
5	031	Braga - IP do Cávado e do Ave
6	040	Bragança - IP de Bragança
7	050	Castelo Branco - Universidade da Beira Interior
8	051	Castelo Branco - IP de Castelo Branco
9	060	Coimbra - Universidade de Coimbra
10	061	Coimbra - IP de Coimbra
11	062	Coimbra - ES de Enfermagem de Coimbra
12	070	Évora - Universidade de Évora
13	080	Faro - Universidade do Algarve

Figura 3.2 Excerto da tabela GAES.

Nome	Descrição
GAES	Código identificador
Descr	Nome descritivo

Tabela 3.2 Descrição dos atributos do GAES.

Os registos de GAES não têm uma relação explícita para a folha de Distrito e, por esse motivo, foi necessário relacionar manualmente o código do distrito com o distrito associado ao nome descritivo dos 37 GAES.

Estabelecimento (Destino dos candidatos)

A folha de cálculo de estabelecimentos do Ensino Superior (Figura 3.3) contém a informação associada ao estabelecimento onde cada candidato ficou colocado, ou seja, destino do candidato.

	A	B	C
1	Estab	EstNomes2	Distrito
2	0100	Universidade dos Açores	19
3	0110	Universidade dos Açores - Angra do Heroísmo	19
4	0130	Universidade dos Açores - Ponta Delgada	19
5	0200	Universidade do Algarve	08
6	0201	Universidade do Algarve - Faculdade de Ciências Humanas e Sociais	08
7	0203	Universidade do Algarve - Faculdade de Ciências e Tecnologia	08
8	0204	Universidade do Algarve - Faculdade de Economia	08
9	0300	Universidade de Aveiro	01
10	0400	Universidade da Beira Interior	05

Figura 3.3 Excerto dos estabelecimento de colocação.

Nome	Descrição
Estab	Código identificador
EstNomes2	Nome descritivo
Distrito	Código do distrito

Tabela 3.3 Descrição dos atributos do estabelecimento de colocação.

Os registos de estabelecimentos de Ensino Superior têm uma relação para a folha de Distrito, através do código de distrito, esta relação permite perceber qual o distrito dos vários estabelecimentos de Ensino Superior.

Distrito

A folha de cálculo de distritos, permite realizar a correspondência do distrito do GAES (origem) e do distrito do estabelecimento de colocação (destino).

	A	B
1	Distrito	Descr
2	01	Aveiro
3	02	Beja
4	03	Braga
5	04	Bragança
6	05	Castelo Branco
7	06	Coimbra
8	07	Évora
9	08	Faro

Nome	Descrição
Distrito	Código identificador
Descr	Nome do distrito

Tabela 3.4 Descrição dos atributos do distrito.

Figura 3.4 Excerto da informação contida na tabela distritos.

3.1.2 Pré-Processamento dos dados

Apesar dos dados se encontrarem num modelo de dados passível de ser diretamente processado, optou-se por realizar uma passo intermédio pelos seguintes motivos:

- Remover e filtrar dados em excesso.
- Otimizar os dados para leitura, facilitando o processamento.
- Validar o conjunto de dados produzidos por terceiros através de um modelo entidade-relacionamento (ER).
- Agregar outras fontes de dados (e.g. população dos distritos).

Validação

Devido ao desconhecimento sobre a integridade dos dados, foi necessário a sua validação através da importação dos ficheiros Excel para uma base de dados relacional MySQL¹⁴, utilizando PHP¹⁵. Este processo permitiu validar as relações dos dados e a sua integridade através de consultas de teste realizadas à base de dados, comparando os resultados com os boletins estatísticos publicados pelo Ministério de Educação. Neste processo foram descobertas algumas inconsistências relativas às faculdades da Universidade de Lisboa (UL) e da Universidade Técnica de Lisboa devido à fusão destas duas instituições no ano de 2014, e a situação pontual da Universidade de Trás-os-Montes e Alto Douro - Escola de Ciências Humanas e Sociais (Chaves) onde foi alterado o código de estabelecimento (Ver Apêndice C).

¹⁴ <https://www.mysql.com/> (consultado em 9-07-2015)

¹⁵ <http://php.net/> (consultado em 9-07-2015)

Agregação de dados adicionais

Para complementar os dados existentes com dados adicionais, foram obtidos os valores populacionais absolutos em cada distrito. Estes dados foram obtidos através do CENSOS 2011¹⁶ publicado pelo Instituto Nacional de Estatística (INE)¹⁷.

Optimização dos dados para processamento

Para facilitar a leitura e compreensão do modelo de dados, foi criado o diagrama de entidades e relações, apresentado na Figura 3.5.

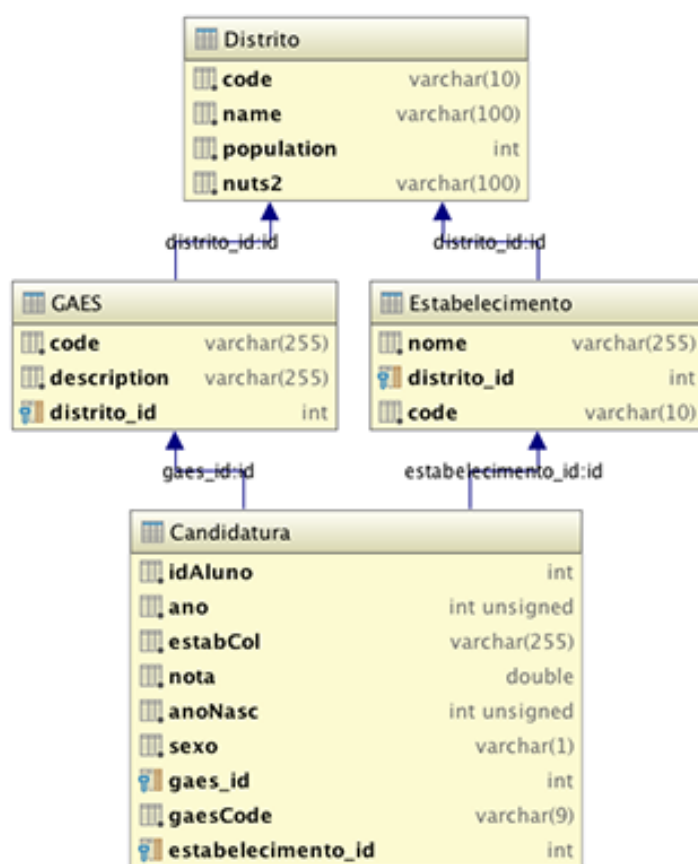


Figura 3.5 - Diagrama ER do modelo de dados.

A Figura 3.6 representa através de um digrama o modelo de dados descritos anteriormente. Depois de relacionados todos os dados é possível determinar a transição dos candidatos dos

¹⁶ <http://censos.ine.pt/> (consultado em 9-07-2015)

¹⁷ https://www.ine.pt/scripts/flex_definitivos/Main.html (consultado em 9-07-2015)

distritos de candidatura para os distritos de colocação. Tomando o exemplo da Figura 3.6, no Distrito A candidatam-se 270 alunos ao Ensino Superior, dos quais 210 ficam colocados no mesmo distrito, 50 transitam para estabelecimentos do Distrito B e 10 para estabelecimentos do Distrito C.

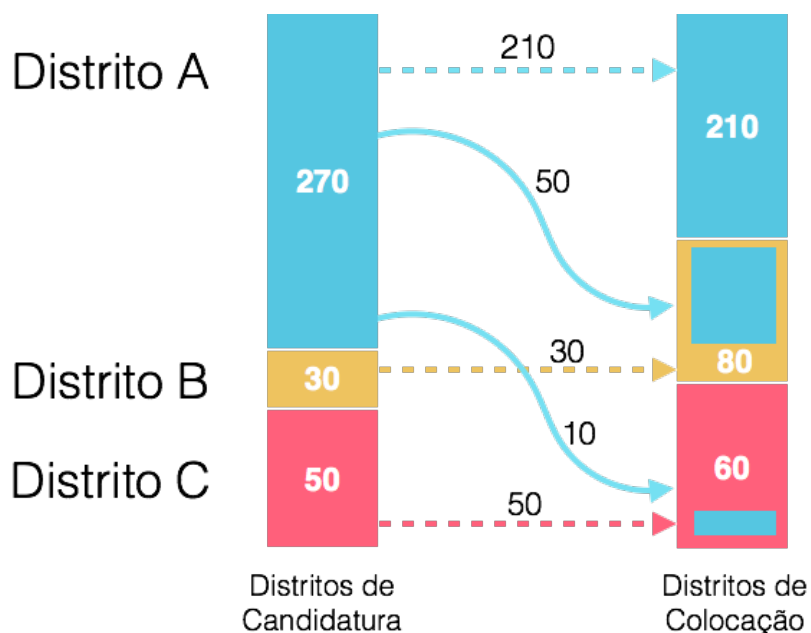


Figura 3.6 Exemplo simplificado da migração de candidatos ao Ensino Superior.

Os dados foram posteriormente exportados para três ficheiros CSV¹⁸ em detalhe na Tabela 3.5. Esta separação em ficheiros diferentes, tem como objectivo reduzir a informação redundante nos dados.

¹⁸ <http://tools.ietf.org/html/rfc4180> (consultado em 9-07-2015)

Ficheiro	Descrição
distritos.csv	Atributos associadas aos distritos CODIGO, NOME, POPULACAO
estabelecimentos.csv	Atributos associadas aos estabelecimentos CODIGO, NOME, CODIGO_DO_DISTRITO
candidaturas.csv	Atributos dos movimentos dos candidatos entre o distrito de origem e distrito de destino. ANO, ANO_NASCIMENTO, SEXO, NOTA, ORIGEM, DESTINO, CODIGO_DO_ESTABELECIMENTO

Tabela 3.5 Descrição do conteúdo dos ficheiros otimizados

Esta optimização permite melhorar a velocidade de leitura e reduzir o tamanho do ficheiro final, passando de um ficheiro inicial com 14,7 MB para três ficheiros que somados totalizam 4,3 MB, reduzindo o espaço necessário em ~70.7%.

3.1.3 Definições gerais

Nas visualizações desenvolvidas são utilizados de forma transversal um conjunto de indicadores, cujo cálculo e definição é descrito na Tabela 3.6. Estes indicadores são utilizados como indicadores numéricos em legendas ou quadros auxiliares ou mapeados em atributos visuais (cores, espessuras e tamanhos) detalhados em cada uma das visualizações.

Sendo v um distrito e V_i^{in} e V_i^{out} o número de candidatos do distrito i colocados em v e o número de candidatos de v colocados em i respectivamente.

Indicador	Definição
Nº total de candidatos de entrada do distrito v	$\sum_{i=1}^{20} V_i^{in}$
Nº total de candidatos de saída do distrito v	$\sum_{i=1}^{20} V_i^{out}$
Saldo total de candidatos do distrito v	$\sum_{i=1}^{20} V_i^{in} - \sum_{i=1}^{20} V_i^{out}$
Rácio: candidatos colocados/total de candidatos desse distrito	$\frac{\sum_{i=1}^{20} V_i^{in}}{\sum_{i=1}^{20} V_i^{out}} * 100$
Percentagem de saída de candidatos do distrito v para o distrito u Sendo $W_{(v \rightarrow u)}$ o número de candidatos de v para u .	$\frac{W_{(v \rightarrow u)}}{\sum_{i=1}^{20} V_i^{out}} * 100$

Tabela 3.6 Indicadores utilizados.

3.2 Representações

As representações desenvolvidas foram alojadas numa plataforma Web (ver Figura 3.7) disponível através no endereço <http://portugal-migration.info>. O trabalho desenvolvido está publicado no plataforma de gestão e versionamento de código da Universidade de Aveiro (CodeUA¹⁹), no endereço <https://code.ua.pt/projects/migration-flow-visualization>. Todos os recursos e implementação técnica encontram-se armazenados no sistema distribuído de versionamento de código GIT, suportado pelo CodeUA.

O website desenvolvido tem como objectivo principal facilitar o acesso às representações, contextualizar o seu âmbito e descrever as várias funcionalidades no mesmo local, auxiliando os utilizadores na compreensão e exploração.

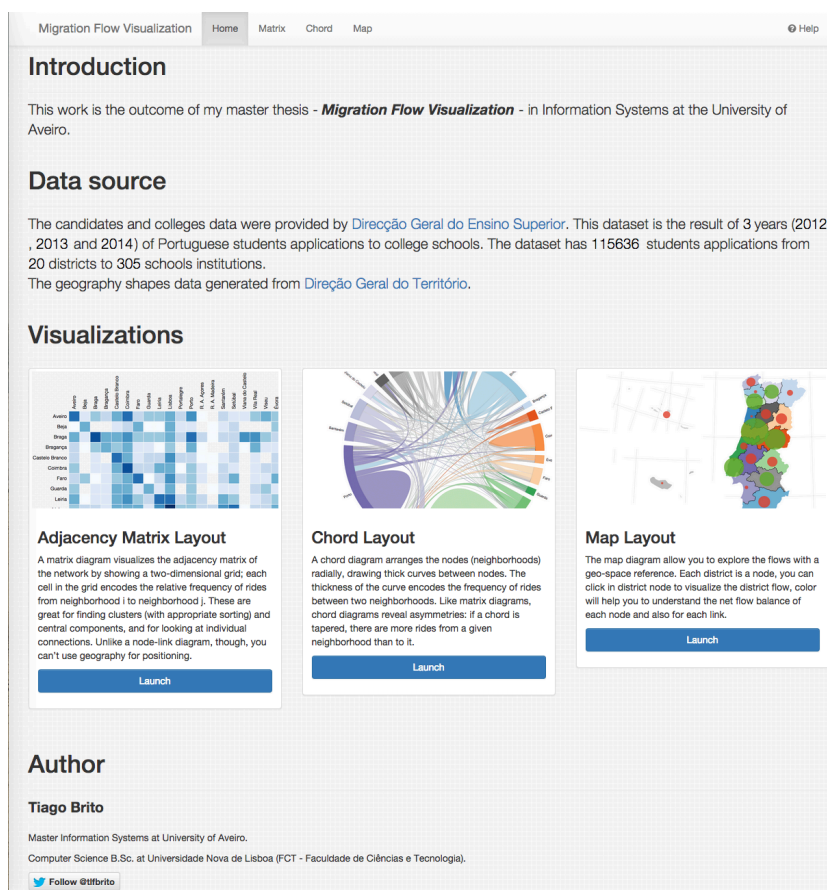


Figura 3.7 Página inicial do website.

¹⁹ <https://code.ua.pt/> (consultado em 12-07-2015)

A página de entrada contém uma introdução a cada uma das representações acompanhada de uma imagem ilustrativa, visível na Figura 3.7. Cada uma das representações pode ser explorada através do seu botão de ação.

3.2.1 Filtros e controlos

As visualizações criadas são acompanhadas por um painel (Figura 3.8) contendo um conjunto de filtros e controlos que permitem filtrar a informação existente e interagir com a representação para alterar ou realçar determinadas características.



Gender: All   Year: 2012 Institution type: All District order: by Name Color: Net flow balance

Figura 3.8 Painel de filtros e controlos da visualização de matriz de adjacências.

Os filtros são comuns a todas as representações e os controlos vão-se alterando de acordo com as características da representação.

Para além dos filtros está também disponível de forma transversal a todas as representações uma *tooltip* (ver Figura 3.10) contendo os vários indicadores englobados na visualização. Este painel, apresenta-se de forma consistente nas três visualizações e é exibido através da interação entre o utilizador e o cursor do rato quando este selecciona um distrito ou um fluxo.

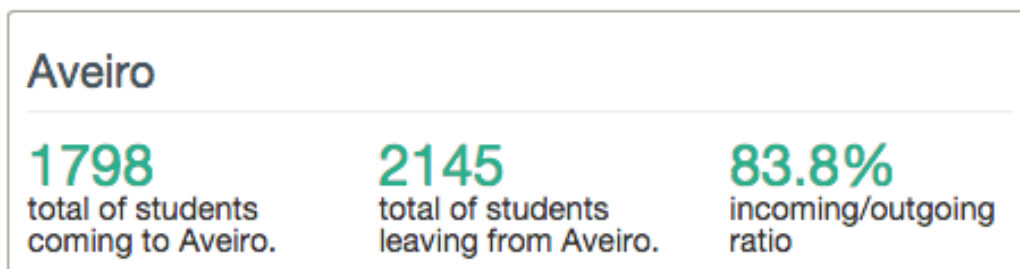


Figura 3.9 Painel de detalhe dos dados do distrito de Aveiro.

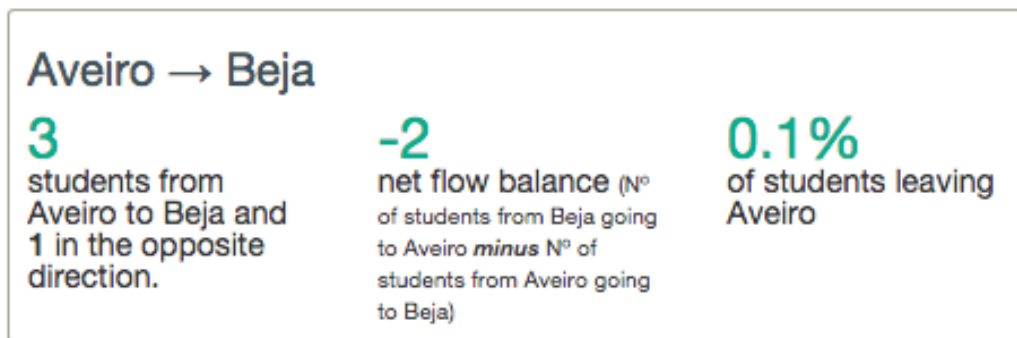


Figura 3.10 Painele de detalhe dos candidatos de Aveiro colocados em Beja.

Este painel apresenta dois tipos de informação distinta:

- **Seleção de Distrito**

Se tiver sido selecionado um distrito, são apresentados os seguintes indicadores relativos a totais do distrito (Figura 3.9):

- **Nº total de candidatos colocados com origem no distrito (saídas)**
- **Nº total de candidatos colocados no distrito (entradas)**
- **Rácio entre Nº total de entradas e Nº total de saídas**

- **Seleção de transição de candidatos (fluxo)**

Se tiver sido selecionado o fluxo, são apresentados indicadores relativos ao fluxo (Figura 3.10):

- **Nº de candidatos** nos dois sentidos.
- Saldo de **candidatos de entrada** menos os **candidatos de saída**.
- Percentagem de candidatos de saída em relação ao **total de candidatos de saída do distrito**.

Para além de macro-indicadores, a seleção do fluxo de candidatos permite-nos também visualizar as instituições do Ensino Superior em que os candidatos foram colocados, como apresenta a Figura 3.11.

Universidade de Aveiro:	256
Universidade de Aveiro - Escola Superior de Saúde de Aveiro:	47
Universidade de Aveiro - Instituto Superior de Contabilidade e Administração de Aveiro:	21
Universidade de Aveiro - Escola Superior de Tecnologia e Gestão de Águeda:	19
Universidade de Aveiro - Escola Superior de Design, Gestão e Tecnologias da Produção de Aveiro-Norte:	9

Figura 3.11 Lista de estabelecimentos com candidatos provenientes do distrito do Porto colocados em Aveiro.

3.2.2 Matriz de adjacências

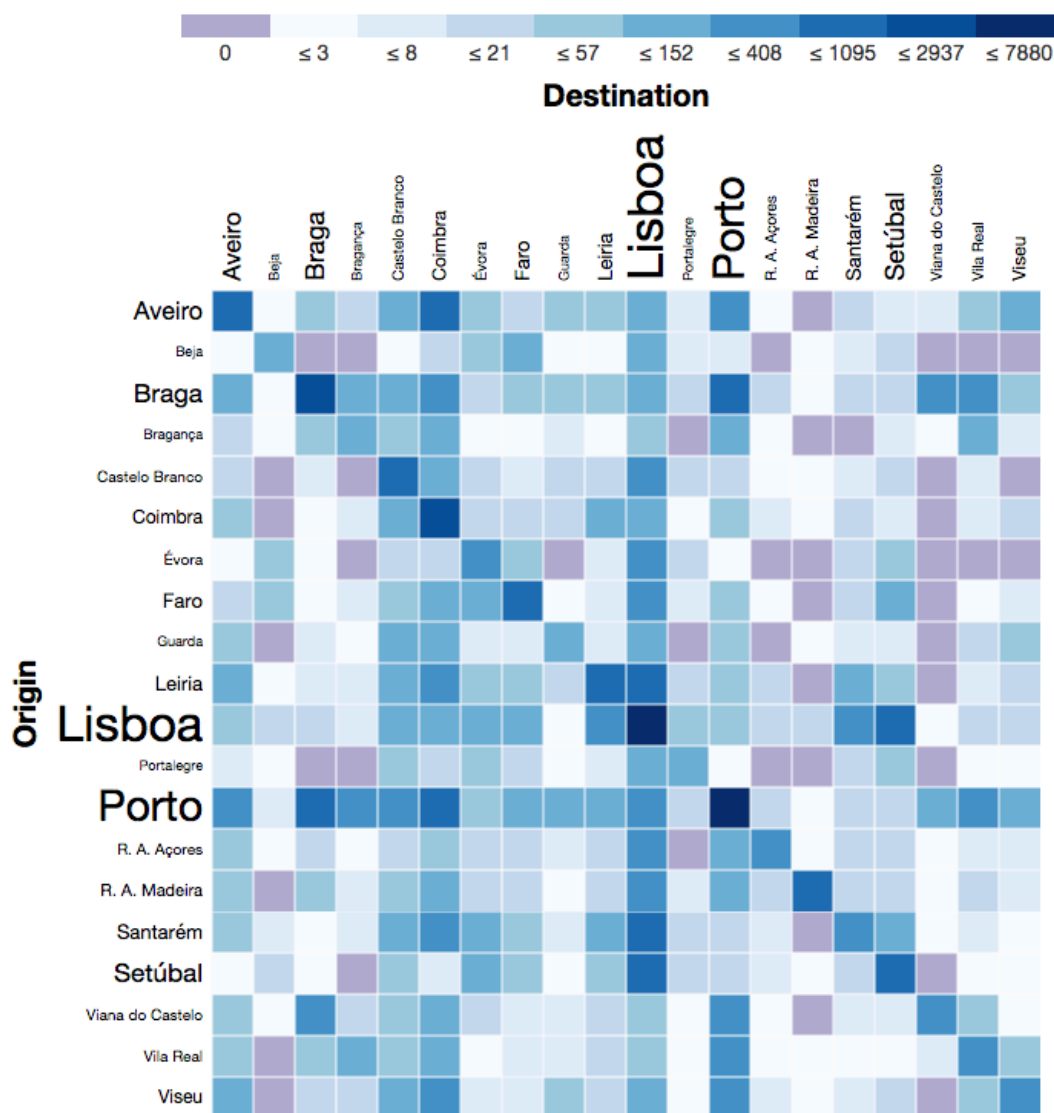


Figura 3.12 Visualização da Matriz de Adjacências.

A estrutura da matriz de adjacências é semelhante a uma tabela de dupla entrada onde do seu lado esquerdo estão os distritos de origem e na parte superior os distritos de destino, as células da matriz representam o número de candidatos dos distritos de origem que foram colocados em estabelecimentos de Ensino Superior do distrito de destino.

Os nomes dos distritos estão ordenados alfabeticamente, de cima para baixo e da esquerda para a direita e o tamanho do texto é proporcional à população do distrito como mostra a Figura 3.13.



Figura 3.13 Legenda dos distritos de destino (tamanho do texto proporcional à população).

Para cada distrito é apresentado um painel informativo, com um conjunto de indicadores que lhe estão associados: total de candidatos colocados no distrito, total de candidatos do distrito colocados noutros distritos e rácio de candidatos colocados por total de candidatos.

Acima da matriz está situada a legenda (Figura 3.14) dos valores utilizados nas células da matriz, esta é composta por uma escala de cores com os valores numéricos associados a cada intervalo. O intervalo mínimo e máximo da escala é calculado e adaptado conforme o conjunto de dados do ano selecionado.

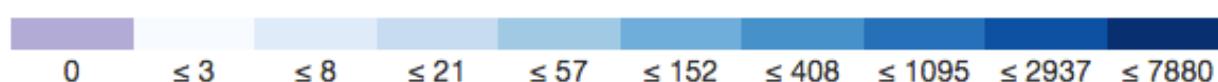


Figura 3.14 Escala de cores e legenda relativa ao número de candidatos.

A cor de cada célula da matriz (Figura 3.14 e Figura 3.17), é configurável através do controlo de cor, por omissão, este controlo encontra-se definido para representar o valor do fluxo migratório (i.e., o número de candidatos colocados), no entanto pode ser alterado pelo utilizador para representar o saldo líquido do fluxo.

No primeiro modo, a cor de cada uma das células representa a quantidade de fluxo de um distrito para o outro, ou seja, o número de candidatos de determinado distrito que se candidatou e foi colocado noutro distrito. Esta cor é tanto mais escura quanto maior for o número de colocados.

A criação da escala de cores para o fluxo de candidatos exigiu esforço adicional, uma vez que os valores dos fluxos são bastante heterogêneos, pois existem fluxos sem qualquer candidato e outros atingindo quase os 8000 candidatos. Esta disparidade de valores compromete a utilização de uma escala linear, que ao ser usada iria apenas enfatizar os fluxos mais elevados.

Para solucionar esta problema, foi criada uma escala logarítmica entre 0 e 100, de forma a restringir a gama de valores originais a um intervalo controlado. Foi criada também uma escala de quantis de forma a distribuir os valores discretos pelas 9 cores do mesmo tom variando apenas na sua crominância, uma vez que estamos a mapear um valor quantitativo sequencial.

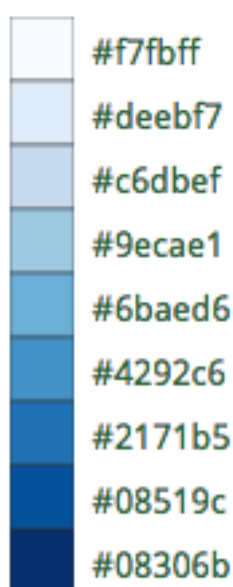


Figura 3.15 Nove tonalidades de azul.

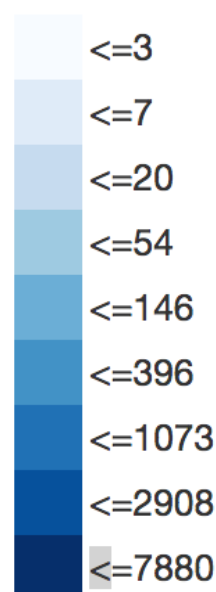


Figura 3.16 Os nove tons com os valores dos quantis.

Já no modo “saldo líquido do fluxo”, a matriz altera a sua escala de cor para quatro valores (zero, negativo, igual ou positivo). Nulo quando não existe qualquer candidato para o par origem-destino, negativo quando o número de candidatos colocados é inferior ao número de candidatos do distrito, positivo caso contrário e igual caso o número de candidatos e o número de colocados do distrito sejam iguais.

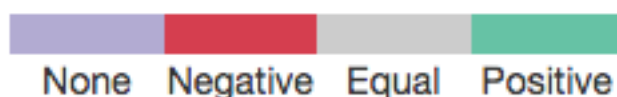


Figura 3.17 Escala do saldo líquido dos fluxos.

Na seleção deste indicador, a escala de cores é substituída de forma a corresponder ao indicador escolhido, passando a apresentar-se como ilustrado na Figura 3.17. Esta alteração desencadeia um conjunto de transições animadas de forma a adaptar a visualização à nova escala, como apresentado na Figura 3.18 e Figura 3.19.

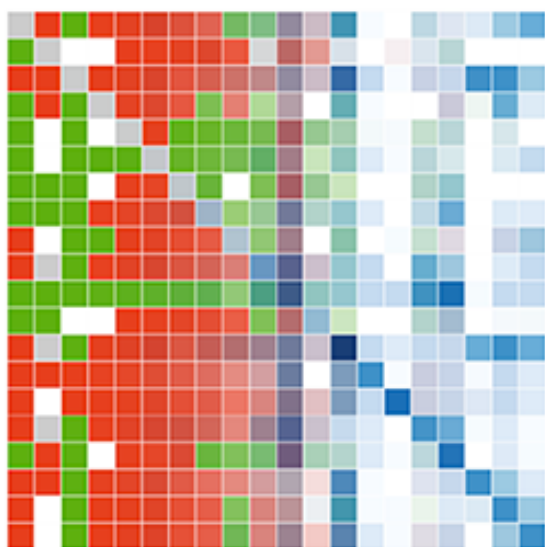


Figura 3.18 Transição entre os dois modos.



Figura 3.19 Matriz no estado de “saldo líquido dos fluxos”.

Uma das características particulares desta representação é a capacidade de ordenar as linhas/colunas da matriz segundo quatro critérios: nome, número total de candidatos do distrito, número total de candidatos colocados no distrito ou população. Esta funcionalidade é especialmente útil para comparar diferenças e destacar padrões, tal como sugerido por (Haroz e Whitney, 2012).

3.2.3 Diagrama cordal

A representação cordal foi a única representação produzida através de um *layout* D3.js predefinido (Chord Layout²⁰), no entanto este é um dos tipos de gráficos mais difíceis de produzir, porque necessita de uma tabela com um formato predefinido e de uma estrutura de dados complementar para apresentar o texto, cores e outros elementos adicionais à representação.

²⁰ <https://github.com/mbostock/d3/wiki/Chord-Layout> (consultado em 9-07-2015)

A representação cordal da Figura 3.20 foi inspirada na representação apresentada na secção 2.4.5, no entanto com algumas diferenças. Tal como na maioria das representações circulares, os arcos representam a quantidade do fluxo, isto é, a quantidade de candidatos que se deslocam de um distrito para outro.

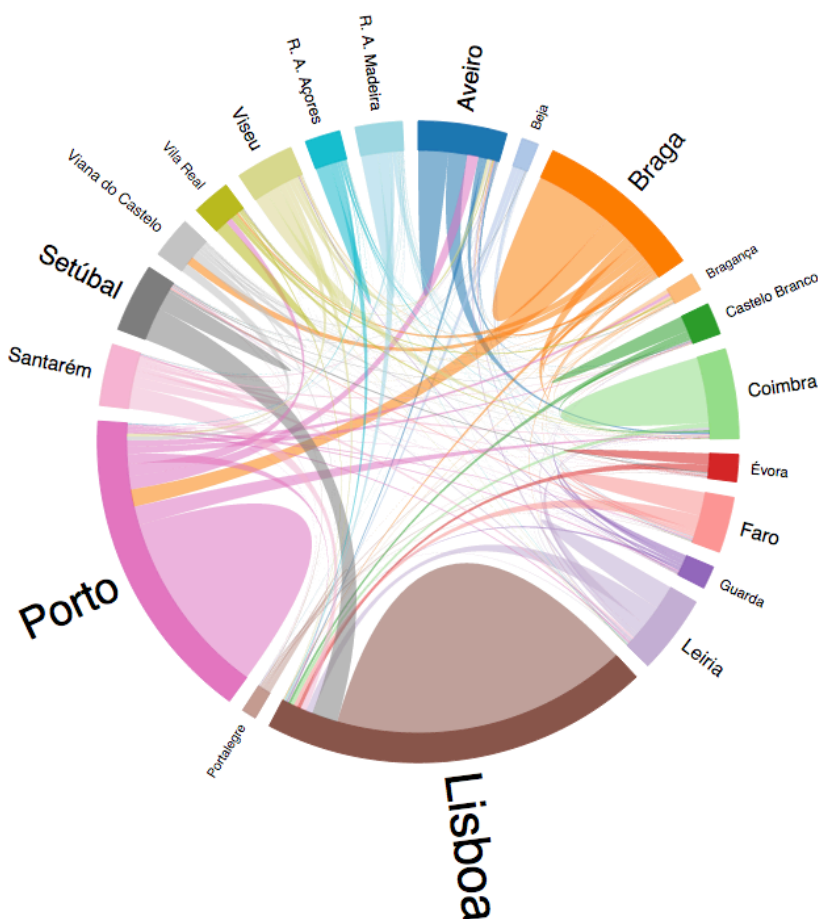


Figura 3.20 Diagrama cordal.

Nesta visualização, ao contrário da representação original os fluxos são todos representados sem qualquer limite mínimo de número de candidatos, no entanto para diminuir a quantidade de arcos visíveis é apenas associado um arco às transições nos dois sentidos. Tomando um exemplo prático significa que a quantidade de candidatos de Aveiro colocados em Lisboa e de Lisboa colocados em Aveiro é representada através do mesmo arco.

A cada distrito está associada uma cor diferente, esta cor é criada através de um conjunto de cores de vinte tons divergentes, uma vez que o objectivo é apenas representar o valor nominal

dos vários distritos e não um valor quantitativo, onde a cada um dos vinte distritos corresponde um tom diferente como sugerido em (Silva *et al.*, 2011).

A cor de cada arco está associada à cor do distrito que origine maior número de colocados, isto é, se 580 candidatos de Santarém forem colocados em Lisboa e por sua vez no sentido inverso apenas 151 candidatos de Lisboa forem colocados em Santarém, a cor do arco (Lisboa-Santarém) será a cor correspondente ao distrito de Santarém.

Outra característica desta visualização é que os arcos de cada distrito estão ordenado por número de candidatos no sentido do ponteiro do relógio. Esta funcionalidade permite identificar rapidamente quais os distritos cujo o maior número de candidatos é colocado no próprio distrito (ex. Lisboa, Porto entre outros) ou não como é o caso de Setúbal e Guarda por exemplo.

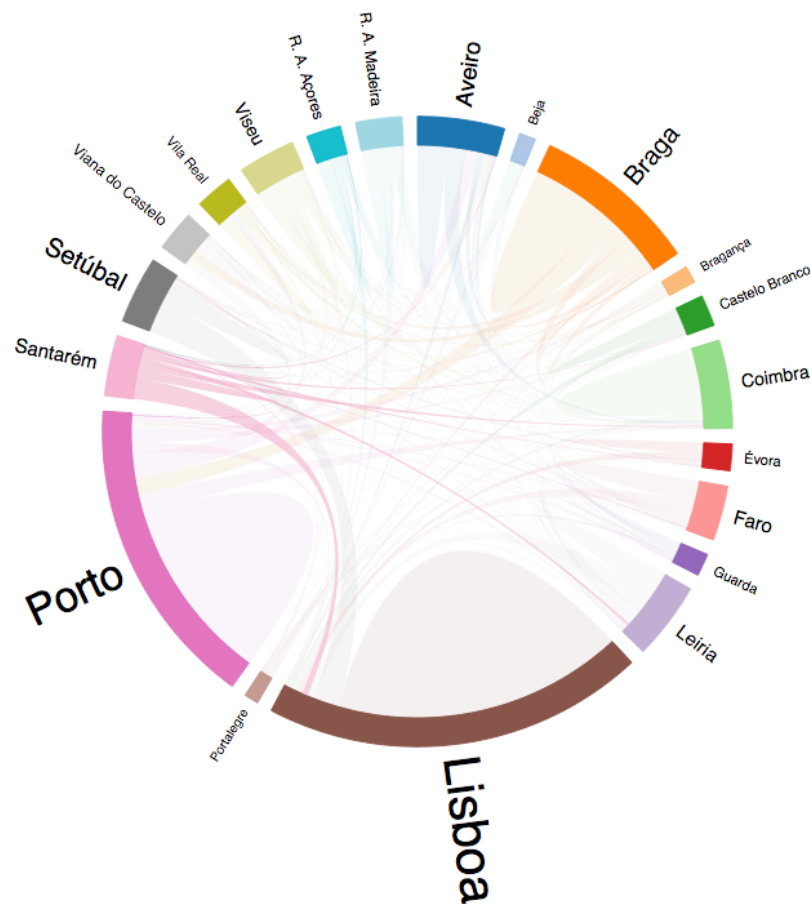


Figura 3.21 Detalhe dos fluxos do distrito de Santarém.

Na periferia do círculo estão representados os vários distritos e estes encontram-se repartidos de acordo com a proporção dos candidatos colocados de cada distrito. O nome de cada distrito está

fixado junto ao arco correspondente e tal como acontece na representação de matriz de adjacências, o tamanho do nome de cada distrito está relacionado com o valor populacional do distrito.

Tal como referido, apesar dos fluxos de candidatos entre os distritos serem representadas por um só arco, o número de candidatos do distrito e candidatos colocados são representados pela espessura das extremidades do arco que une os distritos. Na Figura 3.21 é possível observar que o arco de Santarém para Lisboa é mais espesso na extremidade de Santarém, e mais estreito na extremidade de Lisboa, uma vez que existem mais candidatos de Santarém colocados em Lisboa do que no sentido inverso, justificando a sua cor (cor-de-rosa: Santarém), no entanto pela diferença de espessura é possível ter uma percepção da quantidade de candidatos/colocados em comparação com o resto dos arcos.

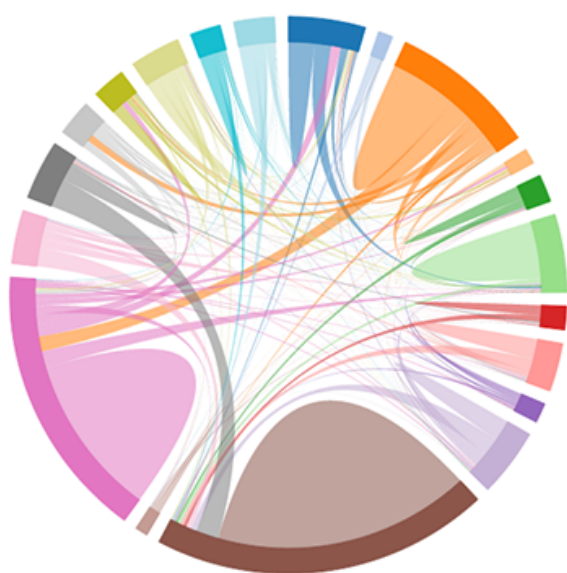


Figura 3.22 Visualização dos candidatos com valores dos filtros por omissão.

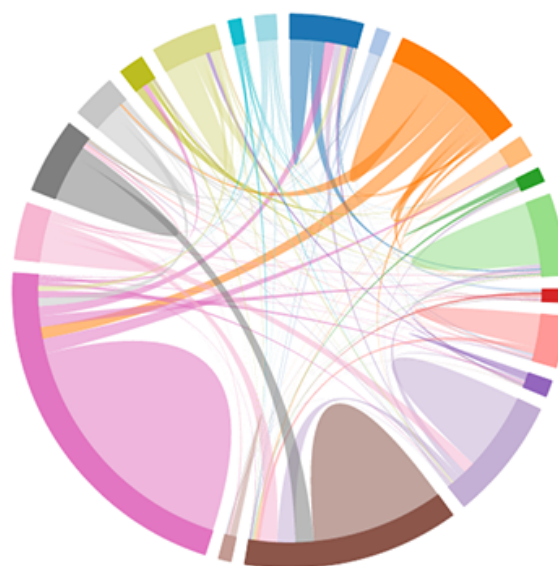


Figura 3.23 Visualização dos candidatos do sexo masculino de 2014 do Ensino Superior Público Politécnico.

A variação dos valores definidos nos controlos de filtragem, permite nesta visualização ter uma percepção melhorada sobre as alterações entre os vários estados selecionados, uma vez que nas implementações tradicionais do *chord layout* a atualização dos dados implica o desenho de um novo gráfico. Nesta implementação sempre que existe alteração do conjunto de dados a representar, a visualização realiza uma interpolação do estado atual (Figura 3.22) para o estado final (Figura 3.23), removendo os arcos desnecessários, adicionando os arcos novos caso existam

e reposicionando os arcos e texto de acordo com os dados atuais, permitindo ao utilizador observar a diferença entre os dois estados e originando um efeito de sucessão do estado atual para o estado subsequente.

3.2.4 Georreferenciação de fluxos

A representação georreferenciada apresentada na Figura 3.24, conjuga a componente geográfica existente nos dados com a sua posição espacial no mapa de Portugal e das suas regiões autónomas.

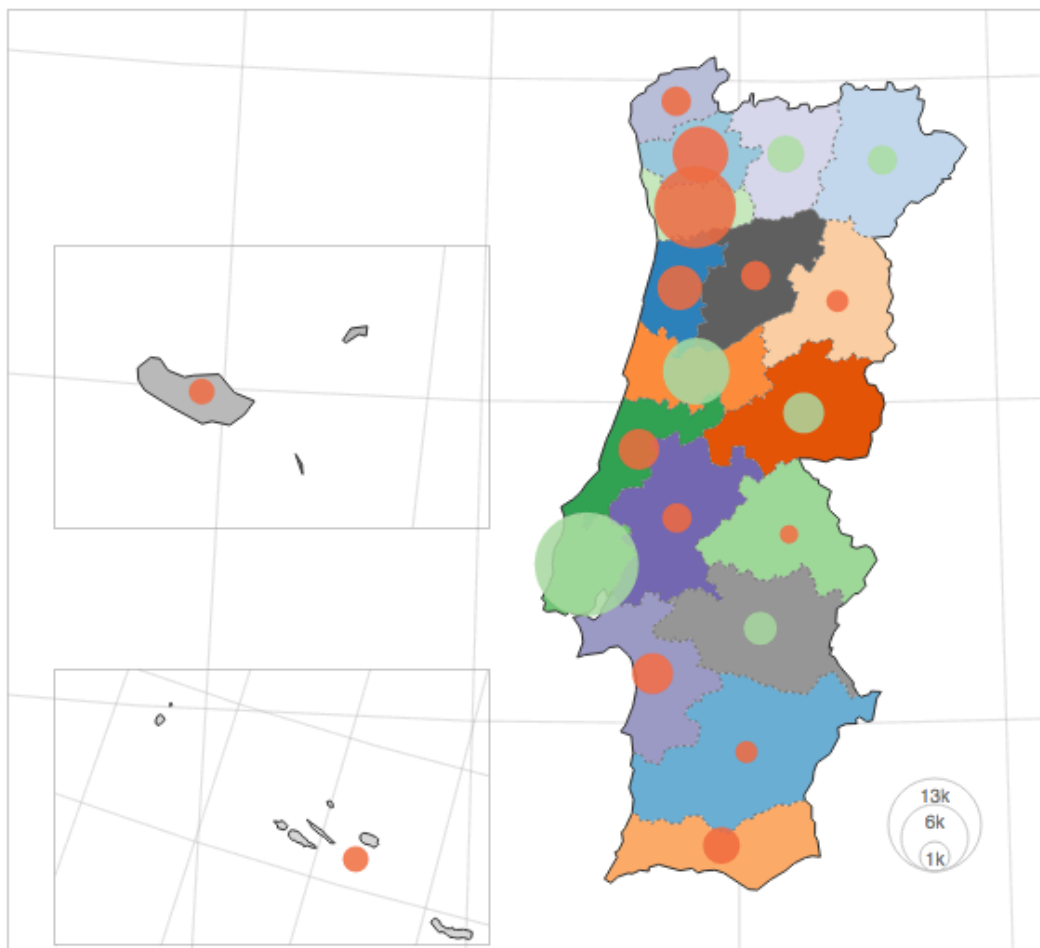


Figura 3.24 Representação de Georreferenciação de fluxos.

Nesta representação, a cor dos distritos é mapeada de dois modos. O primeiro, como mostra a Figura 3.25, com vinte tons divergentes, uma vez que o objectivo é apenas representar o valor

nominal dos vários distritos, para ajudar o utilizador a identificar cada um dos distritos e a sua fronteira e possibilitar uma clara associação da cor ao distrito. É também possível alterar este mapeamento para uma escala de tons azuis, onde é mapeado o total da população no distrito, ou seja, quanto mais escuro for o tom de azul, maior é a população existente nesse distrito (ver Figura 3.26).

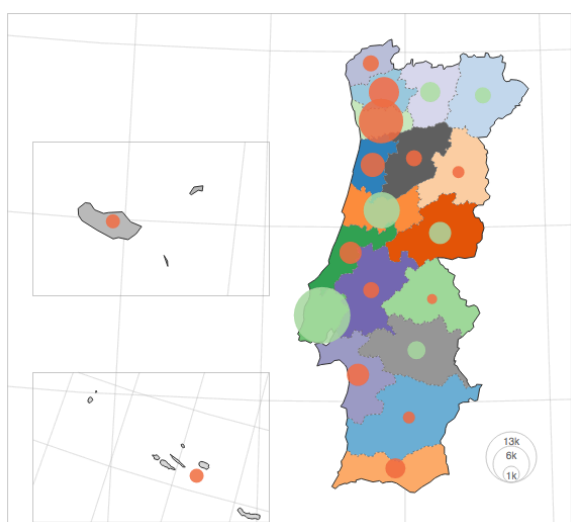


Figura 3.25 Representação dos 20 distritos com 20 cores.

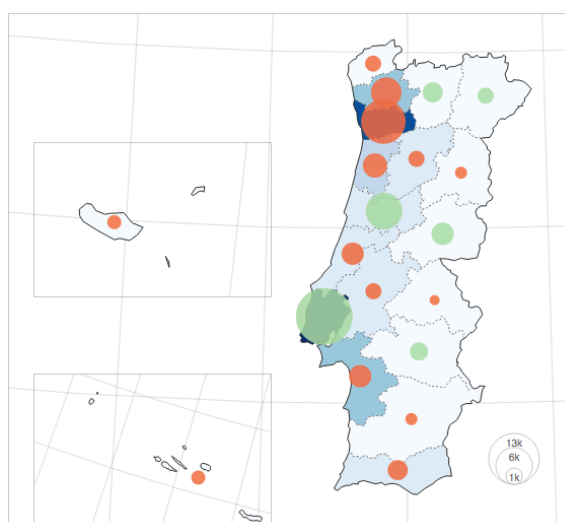


Figura 3.26 Mapeamento da população de cada distrito usando cor.

Representação de Portugal continental e Regiões Autónomas

Apesar da cartografia nacional já existir em várias representações geográficas adaptadas para a Web, a divisão por áreas administrativas e com detalhe elevado é praticamente inexistente, por este motivo, esta visualização exigiu a criação de cartografia específica para ambiente Web. Este desenvolvimento acrescentou uma complexidade adicional na elaboração desta representação, que se encontra descrito em detalhe no Apêndice A.

A representação cartográfica de Portugal continental e regiões autónomas, com as divisões administrativas, foi obtida a partir da Direção-Geral do Território²¹, através da Carta Administrativa Oficial de Portugal (CAOP) em formato *Shapefile*.

Para esta representação geográfica, foi necessária a criação de uma projeção específica, uma vez que ainda não existe nenhuma projeção em D3.js otimizada para Portugal continental e regiões

²¹ <http://www.dgterritorio.pt> (consultado em 9-07-2015)

autónomas. A Figura 3.27 apresenta o resultado da projeção desenvolvida com a cartografia de Portugal continental e regiões autónomas já devidamente modificadas para ambiente Web.



Figura 3.27 - Projeção Albers Portugal criada em D3.js para representar Portugal e regiões autónomas.

A projeção “Albers Portugal” da Figura 3.27 é um caso particular da projeção cónica de Albers que permite a projeção com foco no território Português juntamente com dois *viewports* para incluir as regiões autónomas. Esta projeção foi criada em D3.js no âmbito deste trabalho, pode no entanto ser utilizada fora deste contexto uma vez não depende de quaisquer dados associados.

Representação dos distritos

A cada distrito está associado um círculo de raio variável em função de cada um dos três indicadores disponíveis (Nº total de colocados no distrito, Nº total de candidatos do distrito, rácio dos colocados no distrito), como apresentado na Figura 3.28.

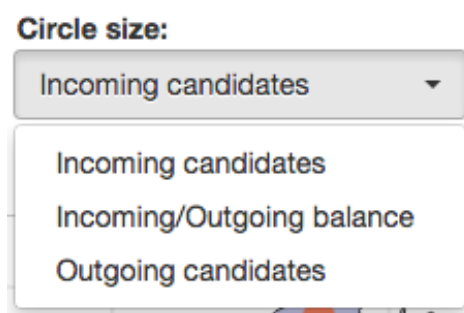


Figura 3.28 Indicadores associados ao tamanho do círculo de cada distrito

Cada círculo está posicionado geograficamente no centróide do distrito a que está associado: isto permite a seleção do distrito em caso de distritos/regiões demasiado pequenos ou regiões dispersas como o caso das ilhas das regiões autónomas. Deste modo evita-se mapear qualquer atributo utilizando a forma/área de cada distrito.

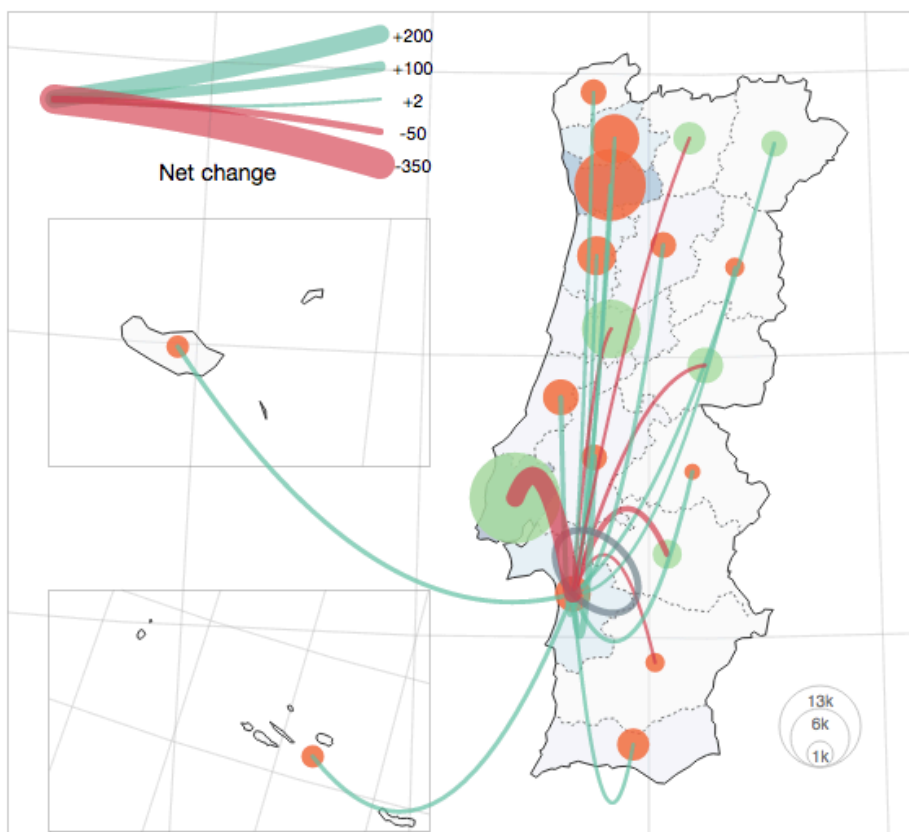


Figura 3.29 Mapa de fluxo após ação do utilizador.

Os círculos são representados através da primitiva SVG *circle*²², a sua cor é definida pela função, onde B_F corresponde ao **Saldo total de candidatos** (ver Tabela 3.6).

$$\text{Cor do círculo} = \begin{cases} \text{Vermelho, se } B_F < 0 \\ \text{Cinzento, se } B_F = 0 \\ \text{Verde - claro, se } B_F > 0 \end{cases}$$

Equação 3.1 Função de cor do círculo

O círculo selecionado é realçado com uma linha negra ao longo do seu perímetro, para facilitar a percepção de seleção, uma vez que os arcos são representados com origem neste círculo/distrito.

²² <http://www.w3.org/TR/SVG/shapes.html#CircleElement>

Representação dos fluxos

Os arcos, que representam os fluxos de candidatos entre os vários distritos, são representações do elemento *Path*²³ SVG que sofrem uma série de alterações de acordo com a tipo de arco a representar.

Os arcos podem ter duas cores, **verde-ciano-claro** caso o fluxo de candidatos de um distrito para outro seja positivo, ou **vermelho** caso contrário, definido pela função da Equação 3.2

Sendo $F_{(v \rightarrow u)}$ o número de estudantes do distrito v para u .

$$\text{Cor do arco} = \begin{cases} \text{Verde – ciano – claro,} & \text{se } F_{(v \rightarrow u)} < F_{(u \rightarrow v)} \\ \text{Cinzento,} & \text{se } v = u \\ \text{Vermelho,} & \text{e } F_{(v \rightarrow u)} > F_{(u \rightarrow v)} \end{cases}$$

Equação 3.2 Cor do arco.

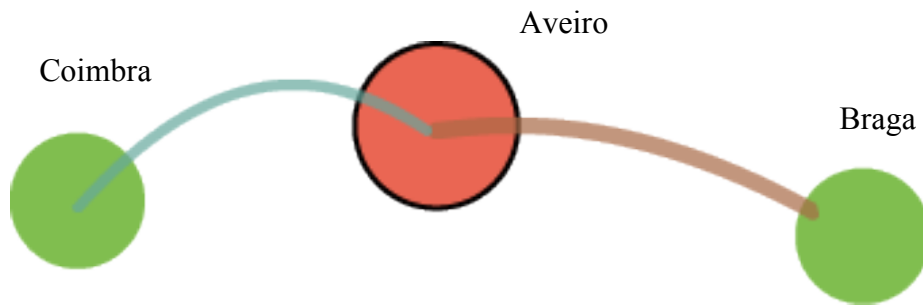


Figura 3.30 Detalhe dos arcos e círculos

Origem	Destino	Nº de candidatos
Aveiro	Coimbra	10
Coimbra	Aveiro	20
Aveiro	Braga	30
Braga	Aveiro	10

Tabela 3.7 Exemplo simplificado dos fluxos entre distritos.

A espessura do arco é proporcional ao valor absoluto da diferença do número de candidatos entre dois distritos, utilizando o exemplo da Figura 3.30 e da Tabela 3.7 Aveiro→Coimbra (10 candidatos) e no sentido inverso Coimbra→Aveiro (20), totalizando o valor absoluto de 10.

²³ <http://www.w3.org/TR/SVGTiny12/paths.html> (consultado em 9-07-2015)

Este mapeamento permite distinguir de forma clara os saldos positivos e negativos de candidatos entre distritos e também o seu valor através da sua espessura.

A transição de candidatos entre os distritos, é representada através de uma curva quadrática de Bézier, definida pelos três pontos de controlo (P_0 , P_1 , P_2), em que P_1 não pertence à curva e P_0 e P_2 são os seus pontos extremos.

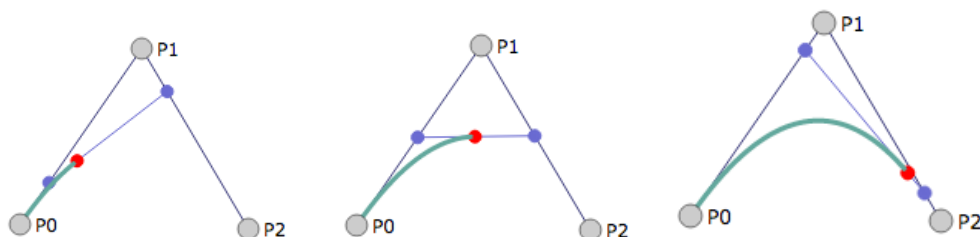


Figura 3.31 Interpolação de uma curva de Bézier de P_0 para P_2 .

Primitiva SVG: `<path d="M598, 124 Q621 99 644 124"></path>`

A instrução anterior ilustra a primitiva necessária para a criação da curva de Bézier Figura 3.31. Inicialmente o cursor de desenho é deslocado (M) para o ponto inicial $P_0(598, 124)$ seguido do ponto de controlo em coordenadas absolutas $P_1(621, 99)$ e terminando com o ponto $P_2(644, 124)$.

Esta técnica é semelhante à utilizada por (Polisciuc *et al.*, 2015) onde é usada uma curva de grau três (i.e., com 4 pontos de controlo) para representar cada arco.

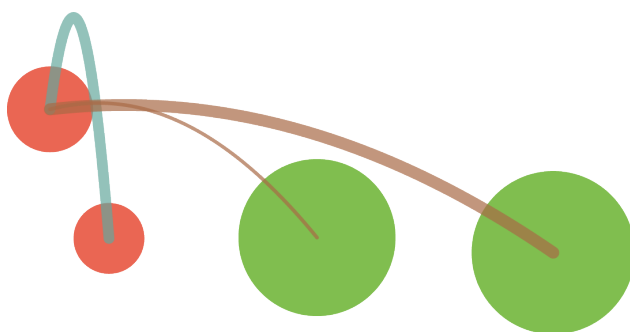


Figura 3.32 Detalhe das curvas de Bézier representando diferentes arcos.

Nesta representação os candidatos que permanecem no próprio distrito, isto é, na situação em que o distrito de candidatura é o mesmo que o distrito de colocação, são representados por um lacete como mostra a Figura 3.33. Cada lacete é desenhado como um arco de elipse. Este arco elíptico é desenhado tendo por base um ponto inicial (x , y), a orientação da elipse é definida por dois raios (rx , ry) e um eixo-rotação- x , que especifica a rotação da elipse em relação ao sistema

de coordenadas SVG actual. O centro (cx , cy) da elipse é calculado automaticamente para satisfazer as restrições impostas pelos outros dois parâmetros, *large-arc-flag* e *sweep-flag* que vão determinar a forma de desenho do arco, como mostra a Figura 3.34.

Neste trabalho é apenas utilizado o caso ilustrado pela combinação da direita em baixo, onde o *large-arc-flag* = 1 e *sweep-flag* = 1 da Figura 3.34.

A Figura 3.33 esquematiza a elipse criada através da seguinte primitiva SVG `<path d="M585,95 A30,20 45,1,1 586,96"></path>` utilizada nos lacetes.

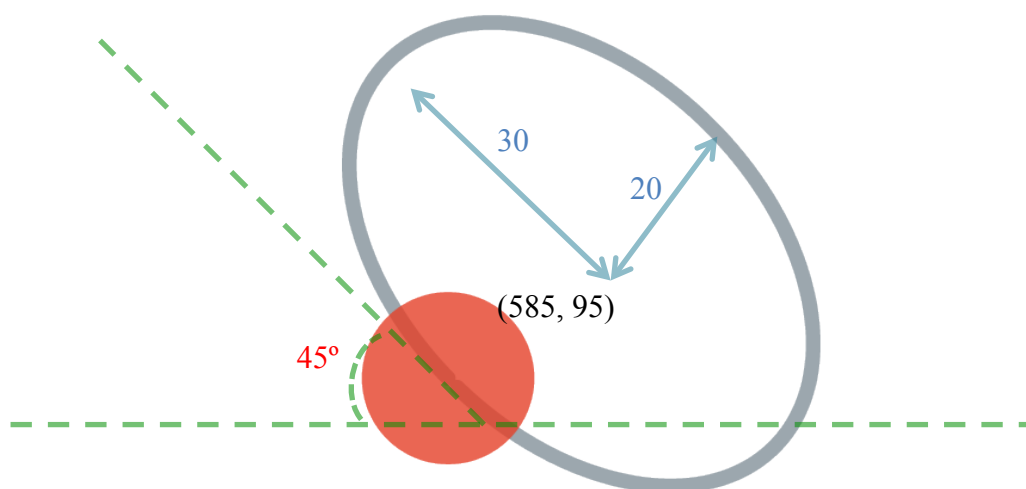


Figura 3.33 – Representação do fluxo para o mesmo distrito.

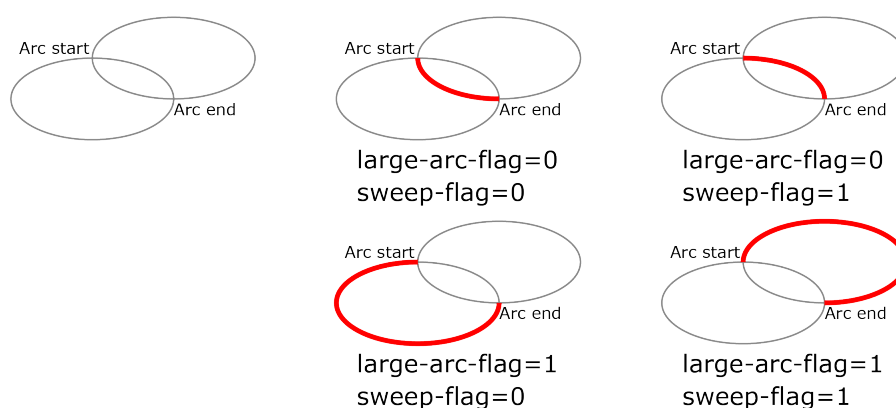


Figura 3.34 Ilustração das várias combinações dos parâmetros para o arco elíptico²⁴.

²⁴ <http://www.w3.org/TR/SVG/paths.html#PathDataEllipticalArcCommands> (consultado em 9-07-2015)

Animação

As animações dos arcos são o resultado de uma interpolação inspirada no exemplo *Stroke Dash Interpolation*²⁵ de Mike Bostock resultando assim no efeito de entrada e saída de um distrito para outro.

A animação dos arcos é realizada de forma distinta para os arcos com saldo de candidatos positivo, dos arcos com saldo negativo, uma vez que os seus pontos de origem e destino são invertidos para realçar o efeito de saída e entrada de candidatos. Esta alteração faz com que ao seleccionar o círculo do distrito, a interpolação dos arcos com fluxo positivo ocorra dos distritos de ligação para o distrito seleccionado e no caso dos arcos de fluxos negativo a interpolação ocorre no sentido inverso, isto é, do distrito seleccionado para os distritos de ligação.

Legendas

Devido à diversidade de elementos visuais e variáveis mapeadas em simultâneo, esta representação contém múltiplas legendas de forma a auxiliar o utilizador na interpretação da informação visual existente.

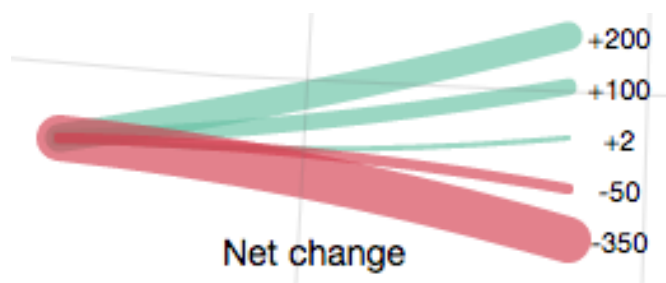


Figura 3.35 Valor do saldo do fluxo entre dois distritos.



Figura 3.36 Número total de candidatos.

A representação é constituída por duas legendas (Figura 3.35 e Figura 3.36), sendo na primeira detalhados os valores associados à espessura dos arcos e a sua cor e na segunda discriminado o número total dos candidatos/colocados.

²⁵ <http://bl.ocks.org/mbostock/5649592> (consultado em 9-07-2015)

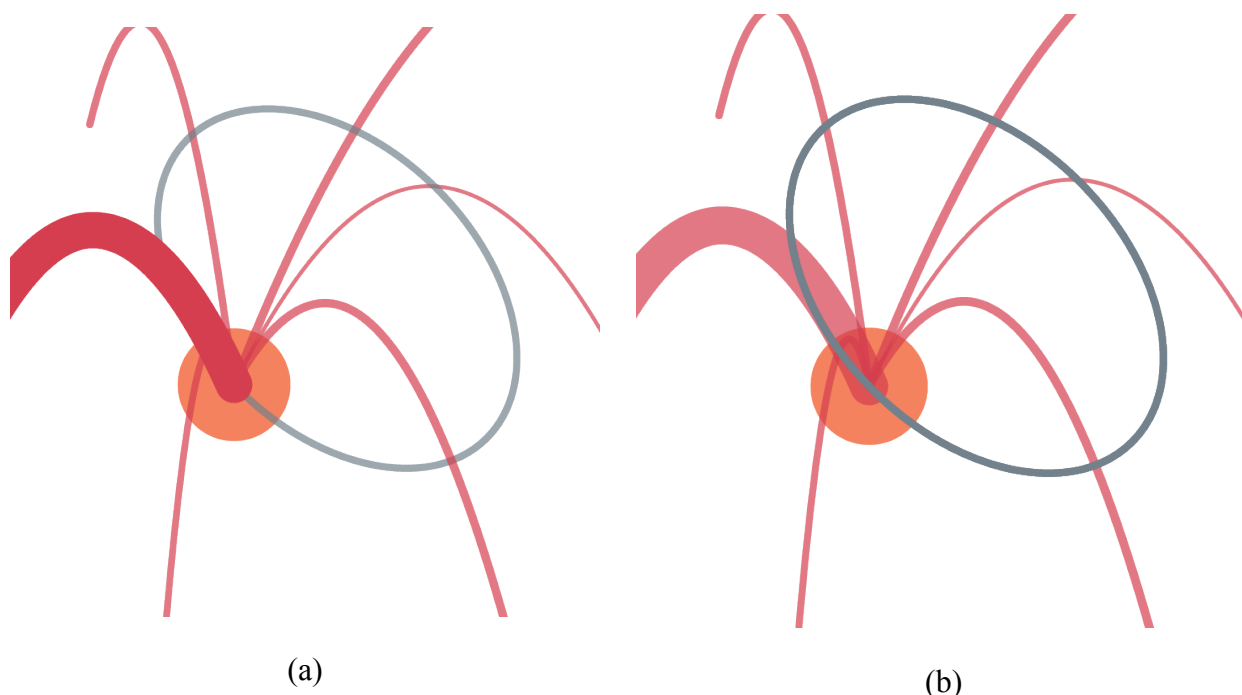


Figura 3.37 Arco selecionado sobreposto a todos os outros.

Esta visualização contém a implementação da sobreposição automática dos arcos de acordo com a seleção por parte do utilizador. Tal como ilustrado na Figura 3.37, à esquerda (a) o arco vermelho encontra-se sobreposto a todos os outros e sem qualquer opacidade definida. Esta disposição é alterada em (b) devido à seleção da elipse cinza, passando esta para o topo, aproximando-se do utilizador, reproduzindo assim o seu propósito natural. Esta funcionalidade tem como objectivo facilitar a seleção e destaque de cada um dos arcos por parte do utilizador.

3.2.5 Ferramentas complementares

Na implementação das representações anteriores, foram utilizadas ferramentas auxiliares em conjunto com a D3.js. Estas ferramentas têm vários propósitos, desde a conversão, processamento, optimização de elementos geográficos, filtragem de grandes quantidade de dados.

Shapefile

*Shapefile*²⁶ é um formato proprietário, desenvolvido pela ESRI²⁷ (Environmental Systems Research Institute) para a descrição de objetos geo-espaciais.

²⁶ <http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf> (consultado em 9-07-2015)

²⁷ <http://www.esri.com/> (consultado em 9-07-2015)

Ogr2ogr

O *ogr2ogr*²⁸ é um utilitário que faz parte do pacote GDAL (*Geospatial Data Abstraction Library*). Este utilitário permite converter *Shapefiles* em *GeoJSON* com a possibilidade de seleção de figuras geométricas com padrão SQL (*Structured Query Language*)²⁹.

GeoJSON

GeoJSON é um formato de definição de figuras geográficas otimizado para a Web uma vez que é subgrupo do formato *JSON*, como é descrito pela sua especificação (Butler *et al.*, 2008).

TopoJSON

*TopoJSON*³⁰ é um formato semelhante ao *GeoJSON*, no entanto, devido à sua estrutura de dados, tem como principal característica a otimização do espaço necessário para a descrição das figuras geográficas, característica essa que se torna especialmente relevante em ambiente Web. O *TopoJSON*³¹ é uma ferramenta criada pelo mesmo criador do D3.js (Mike Bostock) para facilitar e auxiliar o manuseamento (conversão, mapeamentos e outras operações GIS) de objetos geográficos para ambientes Web.



Figura 3.38 – Ilustração dos passos para a geração de TopoJSON simplificado (3) a partir de um Shapefile (1).

Como é ilustrado pela figura anterior, o processo de conversão de uma *Shapefile* em *TopoJSON* é constituído por 2 passos. No primeiro passo (1) a *shapefile* é convertida para *GeoJSON* e as suas formas geográficas podem ser seleccionadas e filtradas de forma a simplificar os dados definitivos. Em (2) o *GeoJSON* é convertido em TopoJSON e ao mesmo tempo os segmentos

²⁸ <http://www.gdal.org/ogr2ogr.html> (consultado em 9-07-2015)

²⁹ <http://en.wikipedia.org/wiki/SQL> (consultado em 9-07-2015)

³⁰ <https://github.com/topojson/topojson-specification> (consultado em 9-07-2015)

³¹ <https://github.com/mbostock/topojson> (consultado em 9-07-2015)

sofrem uma simplificação (visualmente admissível) diminuindo assim espaço necessário para armazenar a figura geométrica original (1). Este processo pode ser analisado com mais pormenor no Apêndice A.

Crossfilter

*Crossfilter*³² é uma biblioteca JavaScript, desenvolvida pela *Square*³³ para explorar grandes conjuntos de dados multidimensionais.

3.2.6 Resultados da exploração interativa dados

Para os anos considerados:

- Setúbal é o único distrito em que a diferença entre os número de candidatos e alunos colocados muda de sinal quando passam a ser apenas do sexo masculino.
- O maior número de candidatos permanece sempre no próprio distrito, exceto em Setúbal.
- Em Lisboa, Porto, Castelo Branco e Coimbra a percentagem de candidatos que ficam no próprio distrito é superior a 50%.
- Setúbal é o único distrito onde o maior fluxo de saída não só não é para o próprio distrito como também é superior a 50% (Lisboa).
- No geral os fluxos mantêm-se sem grandes variações ao longo dos 3 anos.
- Évora e Beja são os distritos cujos os candidatos são colocados num menor número de distritos (Évora: 7 distritos em 2012/2013, Beja: 8 distritos em 2014).
- A R. A. da Madeira (2013-2014) e Viana do Castelo (2012-2014) são as regiões que recebem colocados de um menor conjunto de distritos, um total de dez distritos, de outros dez não recebem qualquer candidato.
- Lisboa tem saldo positivo de candidatos com todos os outros distritos, isto é, todos os outros distritos colocam mais candidatos em Lisboa do que o inverso.

³² <http://square.github.io/crossfilter> (consultado em 9-07-2015)

³³ <https://squareup.com> (consultado em 9-07-2015)

4 Avaliação

À medida que a Visualização de Informação e suas aplicações ganham mais destaque, é importante que novos métodos e representações sejam corretamente validados (Kerren, Stasko e Fekete, 2008). Apesar de importante, este é um tema controverso nesta área, uma vez que tem sido sugerido que as metodologias de avaliação atuais não são suficientemente conclusivas. Isto pode acontecer devido a vários factores: as avaliações de usabilidade são bastante subjetivas, conduzidas com um conjunto de dados pouco expressivo, utilizadores muito homogêneos, ou com tarefas demasiado simplistas, distanciando-se assim das situações reais. As avaliações permitem-nos compreender se o trabalho realizado é realmente útil e perceber se estamos apenas perante uma forma diferente e/ou “simpática” de visualizar os dados, mas que em termos práticos não tem sustentabilidade funcional.

Nas secções deste capítulo, são descritos os dois métodos de avaliação utilizados para avaliar a eficácia e detectar limitações das representações desenvolvidas, e apresentados os seus resultados que permitiram melhorar alguns aspectos e funcionalidades.

4.1 Avaliação Heurística

A avaliação heurística consiste na apreciação de características das visualizações, atendendo a um conjunto de heurísticas realizada por um único avaliador. Desta forma, garante-se uma maior consistência por parte do avaliador e uma avaliação metódica e sistemática.

Assim, e integrado como trabalho na disciplina de Visualização de Informação do Mestrado em Sistemas de Informação, os colegas Manuel Pereira e Ana Gonçalves aplicaram três conjuntos de avaliações heurísticas com base nos modelos de Forsell, Nielsen e Zuk de forma a avaliar a gravidade dos problemas de usabilidade utilizando duas métricas, a aplicabilidade da heurística através de uma escala de zero a cinco e da gravidade do problema detectado numa escala de zero a quatro (Forsell e Johansson, 2010; Nielsen, 1995; Zuk *et al.*, 2006).

Resultados

Os resultados detalhados das diferentes avaliações heurísticas estão descritos nos respectivos relatórios de trabalho (Gonçalves, 2015; Pereira, 2015).

Apresentam-se de seguida as conclusões mais relevantes e as correspondentes modificações efectuadas.

- **Diferente interpretação para daltónicos**

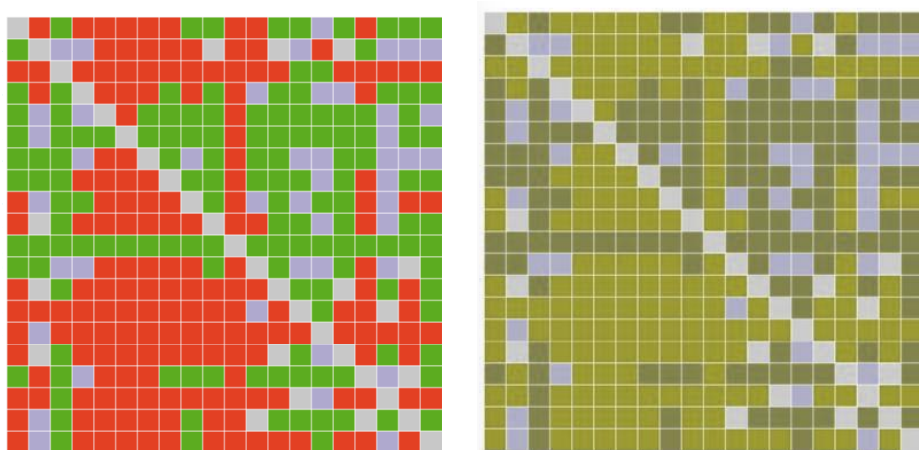


Figura 4.1 Matriz antes da avaliação Heurística.

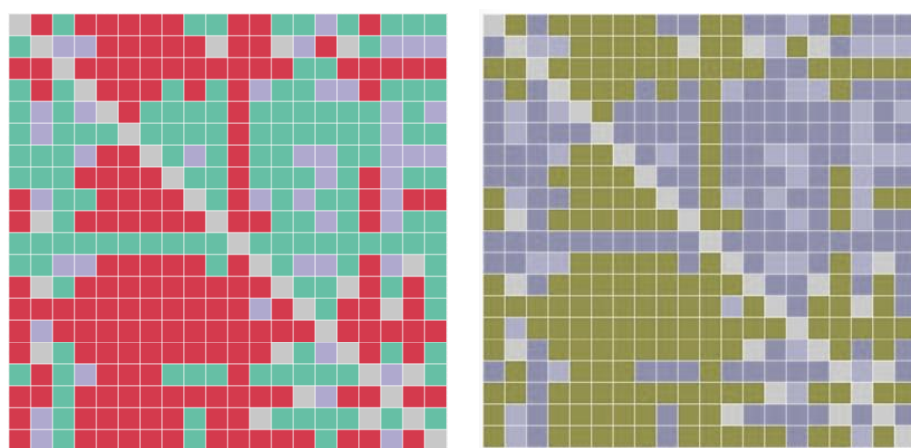


Figura 4.2 Alteração das escalas de cores depois da avaliação heurística.

A Figura 4.1 (antes da correção) e a Figura 4.2 (depois da correção) apresentam a alteração do conjunto de cores não distinguíveis por daltónicos. À esquerda é apresentada a representação original e à direita a representação com as cores modificadas de forma a simular a visão com discromatopsia (daltonismo), revelando assim a perda de informação e impacto que poderá ter

em sujeitos com esta incapacidade, originando interpretações enviesadas. Esta correção permitiu obter uma imagem mais perceptível para utilizadores daltónicos.

- **Estados pouco perceptíveis**



Figura 4.3 Filtro do género antes da avaliação. Figura 4.4 Filtro do género depois do aperfeiçoamento.

A dificuldade de percepção do estado seleccionado pelos controlos, e.g. Figura 4.3, foi outra das situações mais comuns detectadas na avaliação. A solução apresentada na Figura 4.4 consistiu na uniformização da cor dos botões, diferenciando apenas o estado seleccionado e a cor do texto.

- **Ajuda de contexto**

A avaliação destacou também o grande número de termos técnicos (e.g. “*Population*”), cuja compreensão não é intuitiva. Para minimizar este problema, foi adicionado um conjunto de texto descritivo aos controlos e aos valores dos indicadores, para facilitar uma leitura natural. Para além disto, foi também adicionada uma ajuda contextual guiada como ilustrado pela Figura 4.5. Esta ajuda contém uma descrição mais detalhada sobre os controlos e destaca a funcionalidade em contexto.

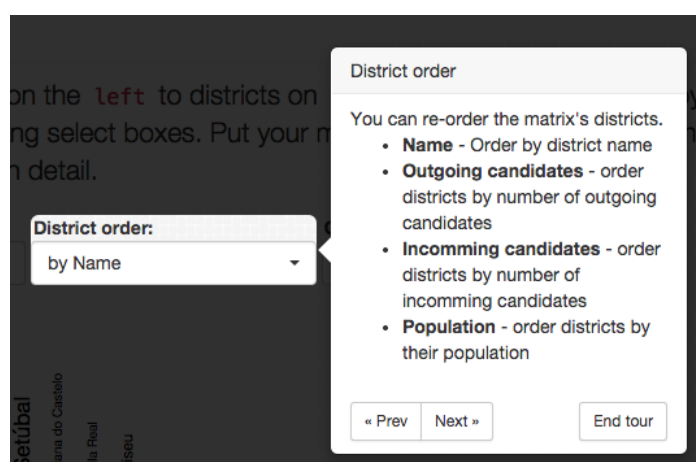


Figura 4.5 Painel de ajuda contextual.

O resultado global das avaliações heurísticas foi de um modo geral bastante construtivo, confirmando a consistência entre as várias representações. Foi destacada positivamente a intenção subentendida em cada uma das representações, de forma a transmitir ao utilizador as diferentes capacidades de cada representação para realizar tarefas específicas.

4.2 Testes de usabilidade

Para além da aplicação metódica das heurísticas, a realização de testes de usabilidade e de experiências controladas continua a ser um dos métodos mais comuns de avaliação em Visualização de Informação (Plaisant, 2004).

Esta metodologia consiste na realização de um conjunto de tarefas com um propósito bem definido. É composta por duas fases distintas, na primeira é realizada uma introdução, a contextualização da ferramenta em teste e uma explicação dos procedimentos práticos do teste. Na segunda fase, os utilizadores têm que realizar o conjunto de tarefas predefinido e avaliar o grau de dificuldade de cada uma das tarefas usando uma escala de Likert.

Neste teste de usabilidade foram criados dois grupos de perguntas. O primeiro é constituído por três secções, uma por cada representação. As tarefas neste grupo são todas equivalentes na sua dificuldade e no objectivo, diferindo apenas na visualização necessária para realizar a tarefa desejada. Estas tarefas têm como principal objectivo introduzir e ambientar os utilizadores às diferentes representações e exibir as principais diferenças entre as mesmas.

No segundo grupo, os utilizadores são confrontados mais uma vez com um conjunto de tarefas, no entanto neste grupo não é indicada qual a visualização que deve ser utilizada para cada tarefa. O utilizador deverá escolher a visualização que lhe pareça mais adequada para efetuar cada tarefa. As tarefas foram planeadas de forma a que os utilizadores tenham liberdade na escolha na representação a utilizar. No entanto, a resolução de algumas tarefas pode ser mais fácil através de uma representação em particular ou até mesmo impossível sem a escolha da representação mais adequada.

Cada utilizador de teste é supervisionado por um observador, que regista o tempo necessário para cada tarefa e o modo como esta é executada, registado também eventuais dúvidas e comentários que surjam durante a realização do teste.

Os questionários foram realizados por 12 utilizadores, reponderam a um total de 21 perguntas numeradas por ordem alfabética, divididas pelos dois grupos descritos em cima. O grupo 1

contém quinze perguntas, cinco por cada representação e o grupo 2 inclui seis perguntas. As questões e os formulários dos observadores são apresentados no Apêndice C.

Resultados

Os resultados dos testes permitem-nos ter uma visão geral sobre a realização e resolução dos testes de usabilidade.

Os resultados seguintes mostram que os utilizadores conseguiram concluir todas as tarefas; no entanto, a tarefa E foi errada por 20% dos utilizadores em todas as representações (ver Figura 4.6).

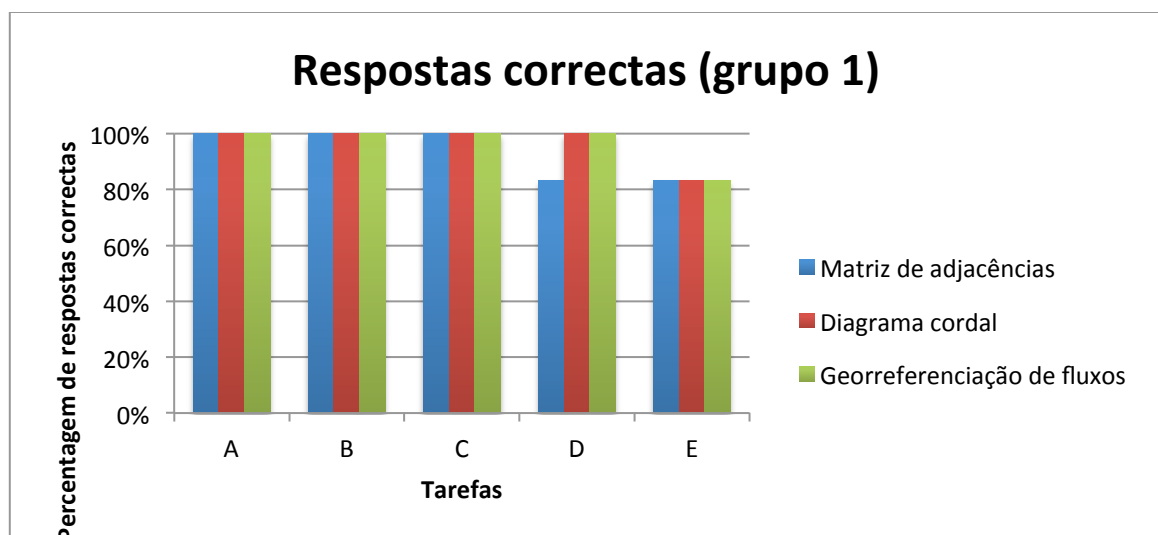


Figura 4.6 Resultado das respostas do Grupo 1.

Esta consistência de respostas erradas à pergunta E poderá indicar um problema de interpretação da tarefa por parte dos utilizadores.

A Figura 4.7 mostra-nos que, apesar de todos os utilizadores terem concluído todas tarefas, a sua resolução foi realizada com tempos bastante distintos. Destacando-se sobretudo os tempos da tarefa B utilizando a representação georreferenciada e da tarefa C, usando a matriz de adjacências.

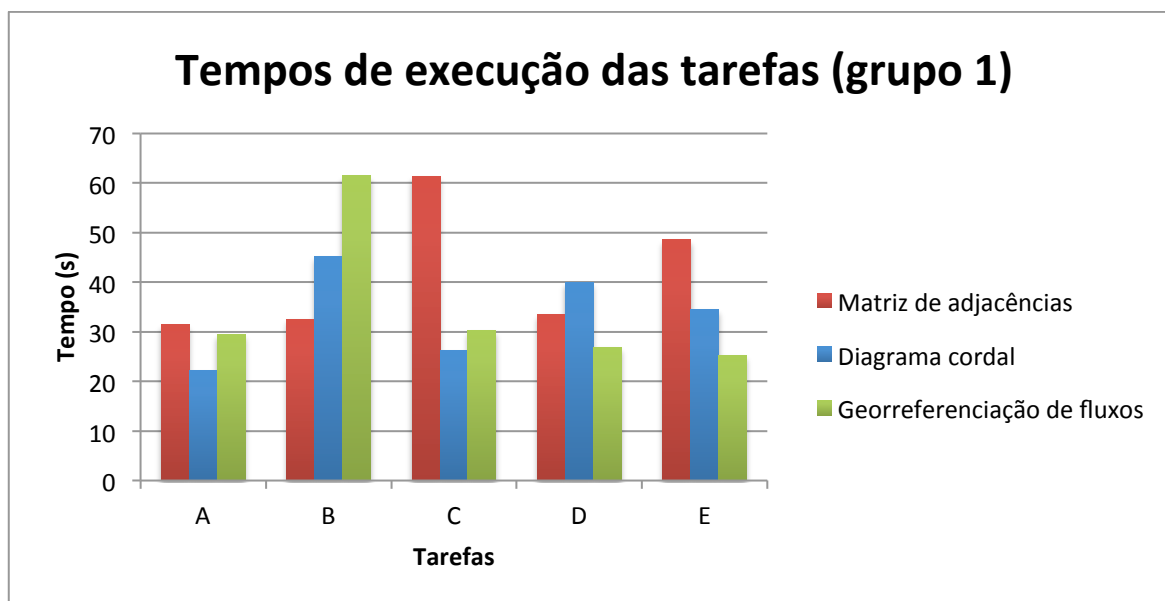


Figura 4.7 Tempos médios de conclusão das tarefas do Grupo 1.

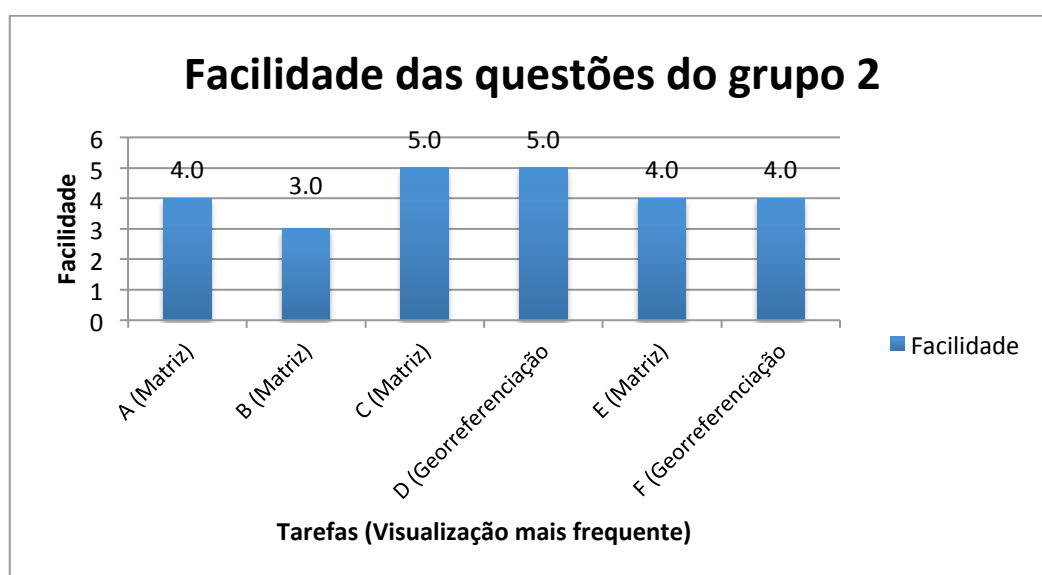


Figura 4.8 Nível de facilidade atribuído pelo utilizador e a visualização mais frequente (Grupo 2).

A Figura 4.8 apresenta-nos uma categorização consistente, por parte dos utilizadores em relação à facilidade de execução de cada uma das tarefas do grupo onde o utilizador era livre de realizar a tarefa através da representação que mais lhe parecesse mais adequada. A representação mais utilizada para cada uma das tarefas é indicada no gráfico em conjunto com a letra da tarefa correspondente. Grande parte dos utilizadores optou por utilizar a representação em matriz de adjacências, excepto na tarefa D e F onde a resolução era apenas possível através do mapa. Estes

dados indicam que, apesar dos utilizadores preferirem a utilização da representação em matriz de adjacências, conseguem identificar claramente a visualização mais correta para resolver tarefas de características georreferenciadas.

Apesar dos resultados anteriores, devido ao reduzido número de utilizadores de teste, os comentários e críticas apontados pelos utilizadores durante a sessão deste teste, são o resultado mais importante na realização deste teste, dentre os quais se destacam pela sua frequência:

- Dificuldade em encontrar os totais dos distritos na representação matricial.
- Dificuldade na leitura das visualizações cordal e georreferenciada devido à quantidade de ligações entre os distritos.
- Na visualização georreferenciada de fluxos, os lacetes não têm o destaque necessário e são difíceis de descobrir.
- Quem não está familiarizado com a geografia de Portugal continental tem dificuldades na utilização da representação georreferenciada, necessitando de procurar o distrito correto através de múltiplas tentativas.
- A legenda da representação matricial passa despercebida em monitores com resoluções reduzidas.
- Dificuldade com a compreensão do significado de alguns indicadores apresentados.

Apesar dos testes de usabilidade supervisionados anteriores se mostrarem bastante proveitosos em termos de comentários, sugestões e críticas, a sua preparação é trabalhosa é difícil angariar utilizadores de teste. Por este motivo muitas das avaliações em artigos mais recentes têm optado pela técnica de *crowdsourcing* cada vez mais popularizada entre a comunidade científica (Heer e Bostock, 2010b; Kittur, Chi e Suh, 2008; Komarov, Reinecke e Gajos, 2013; Kosara e Ziemkiewicz, 2010), esta técnica consiste na angariação de utilizadores para realizar tarefas através de serviços especializados (i.e. Amazon Mechanical Turk³⁴). A utilização deste tipo de testes não faz parte do âmbito inicial desta dissertação, mas devido ao seu interesse será realizado algum trabalho neste sentido.

³⁴ <https://requester.mturk.com> (consultado em 10-07-2015)

5 Conclusões e Trabalho Futuro

Nesta dissertação é estudada a visualização de fluxos migratórios, utilizando como caso de estudo os resultados da primeira fase do concurso nacional de acesso ao Ensino Superior. Diferencia-se das abordagens tradicionais por uma implementação assente apenas em tecnologias baseadas na Web que, em conjunto com uma exploração interativa se revela essencial para melhorar a experiência dos utilizadores simplificando a compreensão dos dados de fluxos migratórios.

Este capítulo irá resumir os objectivos do trabalho, principais contribuições, impacto do trabalho desenvolvido e, finalmente, apresentar algumas sugestões para trabalho futuro.

5.1 Resumo

Este trabalho teve como objectivo principal desenvolver visualizações interativas capazes de representar fluxos migratórios, utilizando um caso de estudo nacional e possibilitar o acesso generalizado ao maior número de pessoas de forma intuitiva e multiplataforma.

Esta dissertação engloba as várias fases do desenvolvimento de visualizações, desde a obtenção dos dados e o seu processamento, desenvolvimento das representações até à avaliação. Para executar este ciclo de desenvolvimento da forma mais adequada foi necessário realizar:

- Uma revisão da literatura existente e das técnicas mais recentes para representações de fluxos migratórios.
- Uma avaliação das visualizações desenvolvidas, através de duas metodologias distintas (avaliação heurística e testes de usabilidade)

Do trabalho desenvolvido resultaram as três visualizações propostas:

- **Representação em matriz de adjacências** – Visualização matricial, onde o número de candidatos é mapeado numa escala de cores e o tamanho do nome dos distritos é proporcional à sua população.
- **Diagrama cordal** – Representação circular do grafo associado às candidaturas, onde os arcos representam as migrações de candidatos.
- **Georreferenciação de fluxos** – Representação das migrações dos candidatos recorrendo aos mapas de Portugal Continental e das Regiões Autónomas.

5.2 Principais contribuições

No trabalho efectuado, destacam-se como principais contribuições, além das três representações desenvolvidas;

- Utilização de um conjunto de dados com informação nacional.
- Criação de uma plataforma para visualização interativa de um grande conjunto de dados, comprovando a maturidade das ferramentas Web disponíveis hoje em dia.
- Desenvolvimento de projeção e mapas do território nacional compatíveis com a Web.
- Validação das diferentes representações através de avaliação de usabilidade.

Outra contribuição indireta foram os resultados obtidos durante do desenvolvimento e exploração da plataforma, compilados na secção 3.2.6.

5.3 Trabalho Futuro

Existem um conjunto de detalhes que poderão ser melhorados, numa próxima fase, tais como:

1. Implementação das sugestões apontadas pelos utilizadores e conclusões dos testes de usabilidade.
2. Massificar os testes de usabilidade através de processos de *crowdsourcing* automatizados, e.g.: *Amazon Mechanical Turk*³⁵, como sugerido por (Kittur *et al.*, 2008; Luther *et al.*, 2015).
3. Continuar com desenvolvimento de mais funcionalidades, utilizando o mesmo conjunto de dados, uma vez que os atributos não foram todos utilizados i.e: nota do candidato.

³⁵ <https://www.mturk.com> (consultado em 9-07-2015)

4. Verificar o comportamento das visualizações através de outro caso de estudo, e.g. poderá ser utilizado como conjunto de dados alternativo (Origem e destino dos estupefacientes em Portugal³⁶ - Relatórios estatísticos anuais da Polícia Judiciária).
5. E por fim, seria também interessante criar uma plataforma independente do caso de estudo.

Espera-se que o trabalho realizado nesta dissertação, cative futuros alunos a desenvolver trabalhos e projetos na área de Visualização de Informação e que isso motive a criação de soluções cada vez mais abrangentes para visualizações de dados migratórios, ajudando os utilizadores a tomarem cada vez melhores decisões através de representações cada vez mais aperfeiçoadas.

³⁶ <http://www.policiajudiciaria.pt> (consultado em 9-07-2015)

Bibliografia

ABEL, G. J.; SANDER, N. - Quantifying global international migration flows. **Science (New York, N.Y.)**. 343:6178 (2014) 1520–2.

ANDRIENKO, N.; ANDRIENKO, G. - Visual analytics of movement: An overview of methods, tools and procedures. **Information Visualization**. 12:1 (2012) 3–24.

BOSTOCK, M.; OGIEVETSKY, V.; HEER, J. - D³: Data-Driven Documents. **IEEE Transactions on Visualization and Computer Graphics**. 17:12 (2011) 2301–2309.

BOYANDIN, I. *et al.* - Using Flow Maps to Explore Migrations Over Time. **Geospatial Visual Analytics Workshop in conjunction with The 13th AGILE International Conference on Geographic Information Science**. 2: (2010) 4.

BRYANT, J. - Visualising Internal Migration Flows. **New Zealand Population Review**,. 37: (2011) 159–171.

BUTLER, H. *et al.* - **The GeoJSON Format Specification** [Em linha], atual. 2008. [Consult. 2 jul. 2015]. Disponível em WWW:URL:<http://geojson.org/geojson-spec.html>.

CHEN, C. - Information visualization. **Wiley Interdisciplinary Reviews: Computational Statistics**. 2:4 (2010) 387–403.

CLEVELAND, W. S.; MCGILL, R. - Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. **Journal of the American Statistical Association**. 79:387 (1984) 531–554.

CORMEN, T. H. *et al.* - **Introduction to Algorithms**. 2. ed. Cambridge : The MIT Press, 2001

DGES - **Concurso Nacional de Acesso: Concurso 2011 em números**

DOANTAM PHAN *et al.* - Flow map layout. Em **IEEE Symposium on Information Visualization, 2005. INFOVIS 2005**. [Em linha]. [S.l.] : IEEE, 2005 Disponível em WWW:URL:<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1532150>.

FEW, S. - Practical Rules for Using Color in Charts. **Visual Business Intelligence Newsletter**. (2008) 1–13.

FORSELL, C.; JOHANSSON, J. - An heuristic set for evaluation in information visualization. **AVI '10 Proceedings of the International Conference on Advanced Visual Interfaces**. 10:3 (2010) 199–206.

FRANCIS, A. M.; SCHNEIDER, J. B. - Using computer graphics to map origin-destination data describing health care delivery systems. **Social Science & Medicine**. 18:5 (1984) 405–420.

- FREEMAN, M. - **Using Storytelling to Effectively Communicate Data**. [S.l.] : Infinite Skills, 2015
- HAROZ, S.; WHITNEY, D. - How capacity limits of attention influence information visualization effectiveness. **IEEE Transactions on Visualization and Computer Graphics**. 18:12 (2012) 2402–2410.
- HEER, J.; BOSTOCK, M. - Declarative language design for interactive visualization. **IEEE transactions on visualization and computer graphics**. 16:6 (2010a) 1149–56.
- HEER, J.; BOSTOCK, M. - Crowdsourcing graphical perception. Em **Proceedings of the 28th international conference on Human factors in computing systems - CHI '10**. New York, New York, USA : ACM Press, 2010b
- HEER, J.; BOSTOCK, M.; OGIEVETSKY, V. - A tour through the visualization zoo. **Communications of the ACM**. 53:6 (2010) 59.
- HEER, J.; SHNEIDERMAN, B. - Interactive Dynamics for Visual Analysis. **Communications of the ACM**. 55:4 (2012) 45–54.
- HO, Q. *et al.* - Implementation of a Flow Map Demonstrator for Analyzing Commuting and Migration Flow Statistics Data. **Procedia - Social and Behavioral Sciences**. 21: (2011) 157–166.
- HONG, S. - Information Visualization : How Information Visualization Aids Human Cognition. **sarahjanehong.github.io**. (2014).
- JENNY, B. - Adaptive Composite Map Projections. **IEEE Transactions on Visualization and Computer Graphics**. 18:12 (2012) 2575–2582.
- KEIM, D.; PANSE, C.; SIPS, M. - Information visualization: Scope, techniques and opportunities for geovisualization. (2004) 1–17.
- KERREN, A.; STASKO, J. T.; FEKETE, J.-D. - **Information Visualization: Human-Centered Issues and Perspectives**. [S.l.] : Springer, 2008
- KITTUR, A.; CHI, E. H.; SUH, B. - Crowdsourcing User Studies With Mechanical Turk. (2008) 453–456.
- KOMAROV, S.; REINECKE, K.; GAJOS, K. Z. - Crowdsourcing performance evaluations of user interfaces. Em **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**
- KOSARA, R.; ZIEMKIEWICZ, C. - Do Mechanical Turks dream of square pie charts? **Proceedings of the 3rd BELIV'10 Workshop on beyond time and errors: novel evaluation methods for Information Visualization - BELIV '10**. (2010) 63–70.
- LIU, S. *et al.* - A survey on information visualization: recent advances and challenges. **The Visual Computer**. (2014) 1373–1393.
- LUTHER, K. *et al.* - Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. Em **Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15**. [S.l.] : ACM Press, 2015
- MEEKS, E. - **D3.js in Action**. 1. ed. Shelter Island, NY : Manning Publications, 2015
- MINARD, C. - **Carte figurative et approximative représentant pour l'année 1858 les émigrants du globe, les pays d'où ils partent et ceux où ils arrivent** [Em linha]. Paris : Charles Joseph Minard, 1862, atual. 1862. [Consult. 3 ago. 2015]. Disponível em WWW:URL:<http://www.loc.gov/resource/g3201e.ct000242/>.

- MUNZNER, T. - **Visualization Analysis and Design**. 1. ed. [S.l.] : A K Peters/CRC Press; Har/Psc edition, 2014
- MURRAY, S. - **Interactive Data Visualization for the Web**. 1. ed. [S.l.] : O'Reilly Media, 2013
- NAGEL, T. *et al.* - Sankey Arcs-Visualizing edge weights in path graphs. **Eurographics Conference on Visualization (EuroVis)**. (2012) 1–5.
- NIELSEN, J. - **How to Conduct a Heuristic Evaluation** [Em linha], atual. 1995. [Consult. 12 jul. 2015]. Disponível em WWW:URL:<http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/>.
- PLAISANT, C. - The challenge of information visualization evaluation. **Proceedings of the working conference on Advanced visual interfaces - AVI '04**. (2004) 109–116.
- POLISCIUC, E. *et al.* - Arc and Swarm-based Representations of Customer's Flows Among Supermarkets. Em **IVAPP 2015 6th International Conference on Information Visualization Theory and Applications, March 2015**
- ROBINSON, A. H. H. - The 1837 maps of Henry Drury Harness. **The Geographical Journal**. 121:4 (1955) 440–450.
- SAKSHUWONG, S.; ANGELI, G. - Adaptive Composite Map Projections in D3. **stanford.edu**. (2013) 7.
- SILVA, S.; SOUSA SANTOS, B.; MADEIRA, J. - Using color in visualization: A survey. **Computers and Graphics (Pergamon)**. 35:2 (2011) 320–333.
- STONE, M. - Choosing Colors for Data Visualization. **Perceptual Edge**. (2006) 1–10.
- TOBLER, W. R. - Experiments In Migration Mapping By Computer. **Geography Department University of California**. Slater (1984).
- TOBLER, W. R. - **Display and analysis of migration tables** [Em linha], atual. 2005. Disponível em WWW:URL:<http://www.geog.ucsb.edu/~tobler/presentations/Movement-talk.pdf>.
- TUFTE, E. R. - **Envisioning Information**. 6. ed. Cheshire : Graphics Press, 1990
- WARE, C. - **Information Visualization: Perception for Design**. 2. ed. [S.l.] : Morgan Kaufmann Publishers, 2004
- WISS, U.; CARR, D.; JONSSON, H. - Evaluating three-dimensional information visualization designs: a case study of three designs. Em **IEEE Conference on Information Visualization. An International Conference on Computer Visualization and Graphics**. [S.l.] : IEEE Comput. Soc, 1998
- ZHU, N. Q. - **Data Visualization with D3.js Cookbook**
- ZUK, T. *et al.* - Heuristics for information visualization evaluation. Em **Proceedings of the 2006 AVI workshop on BEyond time and errors novel evaluation methods for information visualization - BELIV '06** [Em linha]. New York, New York, USA : ACM Press, 2006 Disponível em WWW:URL:<http://portal.acm.org/citation.cfm?doid=1168149.1168162>.

Apêndice A Conversão de dados geográficos

Apêndice A.1 Conversão das *Shapefiles* para TopoJSON

A Figura A.0.1 ilustra o conteúdo cartográfico contido no ficheiro *shapefile* de Portugal continental.

Este ficheiro, apesar de conter a informação necessária apresenta vários desafios. Formato desadequado para a Web, devido ao formato proprietário e ao tamanho (50.6Mb) de cada ficheiro *shapefile*. Infelizmente o nível de divisões administrativas não se encontra agrupado por Distrito mas sim por Freguesia. Foi por isso necessário recorrer à seleção e filtragem das regiões distritais do continente a regiões autónomas.



Figura A.0.1 Representação do conteúdo presente na *shapefile*.

A Figura A.0.2 apresenta os dados discriminados contidos na *shapefile*. Os dados contidos são demasiado detalhados para o objectivo planeado, pelo que foi necessário agrupar as regiões através da técnica ilustrada pela Figura 3.38 e detalhada na secção seguinte.

Freguesia	Município	Distrito	TAA	Area_EA_Ha
União das freguesias de Glória e Vera Cruz	AVEIRO	AVEIRO	ÁREA PRINCIPAL	4532.07
Aradas	AVEIRO	AVEIRO	ÁREA PRINCIPAL	892.77
Ílhavo (São Salvador)	ÍLHAVO	AVEIRO	ÁREA PRINCIPAL	3900.19
Gafanha da Nazaré	ÍLHAVO	AVEIRO	ÁREA PRINCIPAL	1643.59

Figura A.0.2 – Descrição dos metadados associados a polígonos da *shapefile*.

Apêndice A.2 Conversão para GeoJSON

Esta secção pormenoriza os vários passos e respectivos comandos da conversão dos ficheiros shapefile em topoJSON, para tal foram utilizadas as ferramentas Ogr2ogr, GeoJSON e TopoJSON descritas na secção 3.2.5.

Portugal

```
ogr2ogr -overwrite -f GeoJSON -s_srs EPSG:3763 -t_srs EPSG:4326  
continente.geojson continente.shp -dialect sqlite -sql "select  
ST_Union(Geometry), DISTRITO FROM Cont_AAD_CAOP2014 GROUP BY DISTRITO"
```



Madeira

```
ogr2ogr -overwrite -f GeoJSON -s_srs EPSG:2191 -t_srs EPSG:4326 madeira.geojson  
madeira.shp -dialect sqlite -sql "select ST_Union(ST_Buffer(Geometry, 0.0)), ILHA  
as DISTRITO FROM ArqMadeira_AAd_CAOP2014 GROUP BY DISTRITO"
```



No caso da Madeira, por ser um arquipélago tão pequeno, e que só por si contém meta-informação reduzida sobre as várias divisões e subdivisões administrativas, a conversão resulta num *geojson* ligeiramente maior que o *shapefile* original.

Açores Oriental

```
ogr2ogr -overwrite -f GeoJSON -s_srs EPSG:2189 -t_srs EPSG:4326  
acores_oriental.geojson acores_oriental.shp -dialect sqlite -sql "select  
ST_Union(Geometry), ILHA as DISTRITO FROM ArqAcores_GOriental_AAd_CAOP2014 GROUP  
BY DISTRITO"
```



Açores Ocidental

```
ogr2ogr -overwrite -f GeoJSON -s_srs EPSG:2188 -t_srs EPSG:4326
acores_ocidental.geojson acores_ocidental.shp -dialect sqlite -sql "select
ST_Union(Geometry), ILHA as DISTRITO FROM ArqAcores_GOcidental_AAd_CAOP2014 GROUP
BY DISTRITO"
```



Açores Central

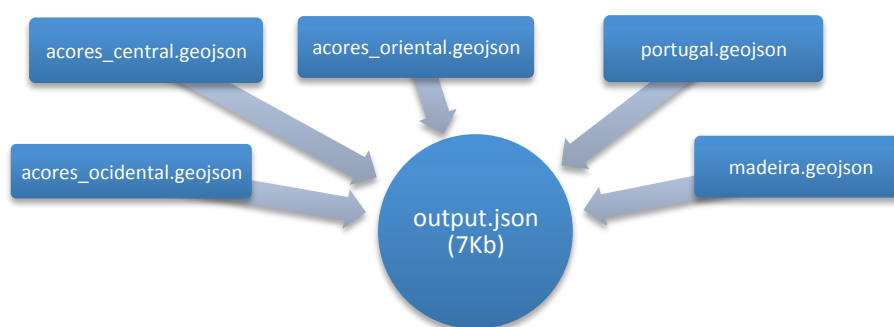
```
ogr2ogr -overwrite -f GeoJSON -s_srs EPSG:2189 -t_srs EPSG:4326
acores_central.geojson acores_central.shp -dialect sqlite -sql "select
ST_Union(Geometry), ILHA as DISTRITO FROM ArqAcores_GCentral_AAD_CAOP2014 GROUP
BY DISTRITO"
```



Conversão e simplificação para TopoJSON

Neste passo os ficheiros topoJSON são agregados num único ficheiro (-o *output.json*), ao mesmo tempo que sofrem uma simplificação³⁷ através da aplicação do algoritmo de Visvalingam³⁸ implementado pelo utilitário de conversão topojson através do parametro “-s 0.000001” de modo a reduzir o seu tamanho de armazenamento sem comprometer os detalhes da cartografia.

```
topojson -p -s 0.000001 -o output.json continente.geojson madeira.geojson
acores_oriental.geojson acores_ocidental.geojson acores_central.geojson
```



³⁷ <http://bost.ocks.org/mike/simplify/> (consultado em 12-07-2015)

³⁸ <https://hydra.hull.ac.uk/resources/hull:8338> (consultado em 12-07-2015)

Apêndice B Migração de dados

Tabela de migração dos códigos das instituições devido à reestruturação do ensino superior nos diferentes anos de 2012-2014

Nome da instituição	Código antigo	Código Novo
Universidade de Lisboa	0700	1500
Universidade de Lisboa - Faculdade de Ciências	0701	1503
Universidade de Lisboa - Faculdade de Direito	0702	1504
Universidade de Lisboa - Faculdade de Farmácia	0703	1505
Universidade de Lisboa - Faculdade de Letras	0704	1506
Universidade de Lisboa - Faculdade de Medicina	0705	1507
Universidade de Lisboa - Instituto de Geografia e Ordenamento do Território	0708	1514
Universidade de Lisboa - Instituto de Educação	0709	1513
Universidade de Lisboa - Faculdade de Psicologia	0710	1511
Universidade Técnica de Lisboa - Faculdade de Medicina Veterinária	0801	1509
Universidade Técnica de Lisboa - Faculdade de Arquitectura	0802	1501
Universidade Técnica de Lisboa - Instituto Superior de Agronomia	0803	1515
Universidade Técnica de Lisboa - Instituto Superior de Ciências Sociais e Políticas	0804	1516
Universidade Técnica de Lisboa - Instituto Superior de Economia e Gestão	0805	1517
Universidade Técnica de Lisboa - Faculdade de Motricidade Humana	0806	1510
Universidade Técnica de Lisboa - Instituto Superior Técnico	0807	1518
Universidade Técnica de Lisboa - Instituto Superior Técnico (campus do Taguspark)	0808	1519
Universidade de Lisboa - Faculdade de Belas-Artes	5302	1502
Universidade de Lisboa - Faculdade de Medicina Dentária	6600	1508
Universidade de Trás-os-Montes e Alto Douro - Escola de Ciências Humanas e Sociais (Chaves)	1205	1202

Apêndice C Testes de Usabilidade: Documentos

Apêndice C.1 Guia de tarefas do Utilizador

Migração dos candidatos ao Ensino Superior público

Sessão de Avaliação

Este documento apresenta uma **lista de tarefas** para serem executadas sobre as visualizações disponíveis em <http://www.portugal-migration.info/>. Este teste de usabilidade destina-se a avaliar o sistema e não o utilizador. Como tal, não se sinta pressionado pelo tempo nem pela necessidade em completar as tarefas com sucesso. Se sentir dificuldades, pode pedir ajuda ou desistir livremente de uma ou mais tarefas.

Execute as tarefas seguintes utilizando apenas o ano de **2012**.

Nº Utilizador: _____

Tarefa 1.1	<p>Para a visualização em Matriz disponível através da entrada Matriz no menu superior.</p> <p>Observando e interagindo com a representação dos fluxos migratórios dos candidatos entre os vários distritos, responda às seguintes questões:</p>																
	<p>A) Quantos alunos de Beja vão para o Porto?</p> <p>Resposta: _____</p>																
	<p>B) Quantos alunos da Madeira ficam na Madeira?</p> <p>Resposta: _____</p>																
	<p>C) Quantos alunos saem no total do distrito de Aveiro?</p> <p>Resposta: _____</p>																
	<p>D) Qual a percentagem da quantidade de alunos de saída do Porto para Bragança?</p> <p>Resposta: _____</p>																
	<p>E) Qual é o distrito que recebe maior quantidade de alunos de Aveiro, excluindo Aveiro?</p> <p>Resposta: _____</p>																
<table border="1"> <tr> <td>Facilidade da Tarefa</td> <td>Nada Fácil</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>Muito Fácil</td> </tr> <tr> <td>1a)</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>		Facilidade da Tarefa	Nada Fácil	1	2	3	4	5	Muito Fácil	1a)							
Facilidade da Tarefa	Nada Fácil	1	2	3	4	5	Muito Fácil										
1a)																	

Tarefa 1.2	Para a representação Cordal disponível através na entrada Chord no menu superior.								
	A) Quantos alunos do Porto vão para o Aveiro?								
	Resposta: _____								
	B) Quantos alunos de Setúbal ficam em Setúbal?								
	Resposta: _____								
	C) Quantos alunos saem no total do distrito da Guarda?								
Resposta: _____									
D) Qual a percentagem da quantidade de alunos de saída de Braga para Coimbra?									
Resposta: _____									
E) Qual é o distrito que recebe maior quantidade de alunos de Braga, excluindo Braga?									
Resposta: _____									
<table border="1"> <tr> <td>Facilidade da Tarefa 1.2)</td> <td>Nada Fácil</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>Muito Fácil</td> </tr> </table>		Facilidade da Tarefa 1.2)	Nada Fácil	1	2	3	4	5	Muito Fácil
Facilidade da Tarefa 1.2)	Nada Fácil	1	2	3	4	5	Muito Fácil		

Tarefa 1.3	Para a representação em Mapa disponível através na entrada Map no menu superior.								
	A) Quantos alunos da R.A. da Madeira vão para o Porto?								
	Resposta: _____								
	B) Quantos alunos de Portalegre ficam em Portalegre?								
	Resposta: _____								
	C) Quantos alunos saem no total do distrito da Guarda?								
Resposta: _____									
D) Qual a percentagem da quantidade de alunos de saída de Vila Real para Aveiro?									
Resposta: _____									
E) Qual é o distrito que recebe maior quantidade de alunos de Faro, excluindo Faro?									
Resposta: _____									
<table border="1"> <tr> <td>Facilidade da Tarefa 1.3)</td> <td>Nada Fácil</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> <td>Muito Fácil</td> </tr> </table>		Facilidade da Tarefa 1.3)	Nada Fácil	1	2	3	4	5	Muito Fácil
Facilidade da Tarefa 1.3)	Nada Fácil	1	2	3	4	5	Muito Fácil		

Tarefa 2	Nesta tarefa deverá escolher livremente a visualização que lhe pareça mais vantajosa para responder a cada uma das questões:	
	A) Quantos distritos não recebem qualquer candidato de Évora?	
	Resposta: _____	
	Facilidade da Tarefa 2a)	<div>Nada Fácil</div> <div> <div>1</div> <div>2</div> <div>3</div> <div>4</div> <div>5</div> </div> <div>Muito Fácil</div>
	B) Qual o distrito de origem e o distrito de destino com maior número de alunos, mas onde o distrito de origem e distrito de destino são distintos?	
	Resposta: _____	
Facilidade da Tarefa 2b)		<div>Nada Fácil</div> <div> <div>1</div> <div>2</div> <div>3</div> <div>4</div> <div>5</div> </div> <div>Muito Fácil</div>
C) Identifique o distrito cujo balanço de alunos (<i>Nº de alunos de entrada – Nº de alunos de saída</i>) seja positivo relativamente a todos os outros distritos?		
Resposta: _____		
Facilidade da Tarefa 2c)		<div>Nada Fácil</div> <div> <div>1</div> <div>2</div> <div>3</div> <div>4</div> <div>5</div> </div> <div>Muito Fácil</div>
D) Concorda com a afirmação “Os candidatos da R.A. da Madeira foram maioritariamente para distritos da faixa litoral” Justifique		
Resposta: _____		
Facilidade da Tarefa 2d)		<div>Nada Fácil</div> <div> <div>1</div> <div>2</div> <div>3</div> <div>4</div> <div>5</div> </div> <div>Muito Fácil</div>
E) Quantos alunos do sexo feminino de Aveiro vão para o Instituto Politécnico de Santarém – Escola Superior Agrária de Santarém?		
Resposta: _____		
Facilidade da Tarefa 2e)		<div>Nada Fácil</div> <div> <div>1</div> <div>2</div> <div>3</div> <div>4</div> <div>5</div> </div> <div>Muito Fácil</div>
F) Concorda com a afirmação “Os dois distritos com maior número de habitantes são também os dois distritos onde são colocados mais alunos” ? Justifique.		
Resposta: _____		
Facilidade da Tarefa 2f)		<div>Nada Fácil</div> <div> <div>1</div> <div>2</div> <div>3</div> <div>4</div> <div>5</div> </div> <div>Muito Fácil</div>

Apêndice C.2 Ficha do Observador

Tarefa	Completoou a tarefa?	Visualização	Sentiu-se perdido / cometeu erros?	Solicitou ajuda?
1 a)	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desistiu	Cordal	<input type="checkbox"/> Sim: _____ <input type="checkbox"/> Não	<input type="checkbox"/> Sim: _____ <input type="checkbox"/> Não
	Observações:			
	Hora Inicio / Fim ____:____/____:____			

Anexos

Tabela de estatísticas da DGES

ACESSO AO ENSINO SUPERIOR 2012-2013

1ª Fase do Concurso Nacional de Acesso de 2012

Mobilidade entre distritos - Distrito de candidatura vs. distrito do estabelecimento de 1ª preferência / colocação

Colocação Candidatura	Aveiro	Beja	Braga	Bragança	Castelo Branco	Coimbra	Évora	Faro	Guarda	Leiria	Lisboa	Portalegre	Porto	Santarém	Setúbal	Viana do Castelo	Vila Real	Viseu	R. A. Açores	R. A. Madeira	Total
Aveiro	799	3	27	17	128	500	24	19	26	48	107	8	297	14	7	3	29	86	1	0	2143
Beja	1	111	0	0	3	10	37	67	2	2	142	7	4	4	17	0	0		0	1	408
Braga	100	3	2073	105	119	265	15	31	23	44	129	12	669	9	13	222	231	33	9	1	4106
Bragança	11	1	22	113	26	60	3	3	4	1	37	0	77	0	8	1	55	4	1	0	427
Castelo Branco	13	0	6	0	412	96	10	4	9	16	157	10	9	7	15	0	4	0	1	1	770
Coimbra	54	0	3	4	57	1745	10	14	14	65	128	2	27	15	4	0	7	18	4	1	2172
Évora	3	27	1	0	16	13	308	29	0	4	194	10	2	9	37	0	0	0	0	0	653
Faro	14	31	2	4	28	77	64	631	1	8	374	8	23	15	57	0	3	4	3	0	1347
Guarda	37	0	5	3	116	142	4	4	79	5	72	0	35	4	4	0	10	31	0	1	552
Leiria	72	2	6	6	81	295	40	26	14	655	533	12	35	59	47	0	5	11	10	0	1909
Lisboa	24	18	9	6	66	91	104	97	3	178	7873	24	28	151	691	1	9	10	16	14	9413
Portalegre	5	3	0	0	34	19	53	13	2	7	116	65	2	9	22		2	1	0	0	353
Porto	352	3	459	171	192	501	26	58	56	68	263	10	4977	19	17	102	325	78	13	3	7693
Santarém	51	4	1	5	74	169	74	22	6	112	580	17	19	276	65	1	4	1	7	0	1488
Setúbal	2	18	2	0	23	8	72	50	3	24	885	9	9	19	522	0	3	1	5	1	1656
Viana do Castelo	54	1	221	18	30	93	9	7	7	9	50	1	236	4	8	216	39	3	2	0	1008
Vila Real	32	0	54	76	31	102	3	6	6	13	38	2	166	2	3	5	347	40	1	1	928
Viseu	125	0	17	17	106	357	6	6	28	20	134	2	165	6	14	0	49	345	5	1	1403
R. A. Açores	24	2	16	3	18	50	13	10	4	15	192	0	72	9	20	2	8	4	388	2	852
R. A. Madeira	25	0	26	7	29	95	12	17	2	16	294	5	100	9	20	3	16	4	12	442	1134
Total	1798	227	2950	555	1589	4688	887	1114	289	1310	12298	204	6952	640	1591	556	1146	674	478	469	40415

Tabela oficial DGES de estatística da migração dos estudantes do ensino superior em 2012³⁹

³⁹ <http://www.dges.mctes.pt/DGES/pt/Estudantes/Acesso/Estatisticas/EstudosEstatisticas/>

Avaliações Heurísticas

Avaliação de Manuel Pereira

departamento de electrónica, telecomunicações e informática	universidade de aveiro		theoria poiesis praxis
Information Systems Master Degree			

Heuristic Evaluation

Title: Information Visualization Report
Author: 43492 - Manuel Pereira
Topic: Heuristic Evaluation
Date: March 2015

Work carried out under the discipline
of
Information Visualization
on
Information Systems Master Degree

Responsible Lecturer
Beatriz Sousa Santos



Heuristic Evaluation

INDEX

1. ABSTRACT	3
2. INTRODUCTION	4
3. METHOD	4
4. NIELSEN'S HEURISTICS	5
5. ZUK'S HEURISTICS	6
6. FORSELL'S HEURISTICS	7
7. CASE STUDY	8
8. ADJACENCY MATRIX VISUALIZATION	9
8.1 NIELSEN'S HEURISTIC EVALUATION	9
8.2 ZUK'S HEURISTIC EVALUATION	10
8.3 FORSELL'S HEURISTIC EVALUATION	11
9. CHORD VISUALIZATION	12
9.1 NIELSEN'S HEURISTIC EVALUATION	12
9.2 ZUK'S HEURISTIC EVALUATION	13
9.3 FORSELL'S HEURISTIC EVALUATION	14
10. MAP VISUALIZATION	15
10.1 NIELSEN'S HEURISTIC EVALUATION	15
10.2 ZUK'S HEURISTIC EVALUATION	16
10.3 FORSELL'S HEURISTIC EVALUATION	17
11. CONCLUSIONS	18
12. REFERENCES	18

ILLUSTRATIONS INDEX

ILLUSTRATION 1 – VISUALIZATIONS OVERVIEW ON THE MAIN SITE PAGE	8
ILLUSTRATION 2 – ADJACENCY MATRIX VISUALIZATION	9
ILLUSTRATION 3 – ADJACENCY MATRIX VISUALIZATION WITH PROTANOPIA FILTER	10
ILLUSTRATION 4 – CHORD VISUALIZATION	12
ILLUSTRATION 5 – CHORD VISUALIZATION WITH PROTANOPIA FILTER	13
ILLUSTRATION 6 – MAP VISUALIZATION	15
ILLUSTRATION 7 – MAP VISUALIZATION WITH PROTANOPIA FILTER	16



1. Abstract

Information systems are used to collect, store, process and distribute information generated by all types of users. Frequently these systems are used to support decision making tasks therefore making the Information Visualization (InfoVis) techniques important. Experts recommend the use of heuristics to facilitate the selection process of Information Visualization techniques to be applied in any context.

The goal of this academic report was to analyze the applicability of three well known sets of heuristics in Tiago Brito Master Theses on "Migration Flow Visualization" in Information System discipline.

The sets of heuristics selected for this evaluation where Nielsen's, Zuk's and Forsell's heuristics.



2. Introduction

Usability heuristics were introduced and developed by Nielsen and Mack [1994], focus on the usability of the interface and are design to be applied to any application, therefore are useful of use to Information Visualization (InfoVis). Despite that there are areas of Information Visualization that these heuristics at best pass over. In response Zuk and Carpendale [2006] consider developing heuristics based on the advice of experts such as design advice, semiotic considerations and/or research in cognitive and perceptual science. More recently Forsell and Johansson [2010] presented a new set of 10 heuristics out of 63 heuristics (from 6 earlier published heuristic sets). This new set is especially tailored to the evaluation of common and important usability problems in Information Visualization techniques.

3. Method

The heuristic evaluation was based on three sets of heuristics, Nielsen, Zuk and Forsell set's of heuristics described in dedicated topics of this report.

It is commonly assumed that three to five expert evaluators are sufficient for a heuristic usability evaluation [Holzinger, 2005] [1]. However we are going to conduct a study with one evaluator, with reserved knowledge about usability principles.

To ensure impartial evaluation, the evaluations were performed isolated by the evaluator.

The current recommended approach to heuristics is that evaluators apply the heuristics in a two pass method. The first pass is done to gain overview of the system and the second to evaluate with more detail each interface component with each heuristic.

In this evaluation the applicability rating scale was the same 0-5 scale that was used by Nielsen:

0 = the heuristic does not explain the usability problem at all.

1 = the heuristic may superficially address some aspects.

2 = the heuristic explains a small part of the problem, but there are major aspects of the problem that are not explained.

3 = the heuristic explains a major part of the problem, but there are major aspects of the problem that are not explained.

4 = the heuristic gives a fairly complete explanation, but there is still more to the problem than is explained.

5 = the heuristic gives a complete explanation.

The severity rating scale was also the same 0-4 scale that was used by Nielsen for severity of usability problems:

0 = I don't agree that this is a usability problem at all.

1 = Cosmetic problem only: need not be fixed unless extra time is available on project.

2 = Minor usability problem: fixing this should be given low priority.

3 = Major usability problem: important to fix, so should be given high priority.

4 = Usability catastrophe: imperative to fix this before product can be released.



4. Nielsen's heuristics

The first set of heuristics to use was the ten usability heuristics from Nielsen applied to user interface design. [2]

N1 - Visibility of system status

The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

N2 - Match between system and the real world

The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

N3 - User control and freedom

Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

N4 - Consistency and standards

Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

N5 - Error prevention

Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

N6 - Recognition rather than recall

Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

(Read full article on recognition vs. recall in UX.)

N7 - Flexibility and efficiency of use

Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

N8 - Aesthetic and minimalist design

Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

N9 - Help users recognize, diagnose, and recover from errors

Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

N10 - Help and documentation

Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.



5. Zuk's heuristics

The second set of heuristics to use was the ten heuristics from Zuk. [\[3\]](#)

Z1 - Ensure visual variable has sufficient length [\[3\]\[25\]\[26\]](#)

Z2 - Don't expect a reading order from color [\[3\]\[25\]\[26\]](#)

Z3 - Color perception varies with size of colored item [\[25\]\[3\]\[26\]](#)

Z4 - Local contrast affects color & gray perception [\[25\]\[26\]](#)

Z5 - Consider people with color blindness [\[25\]\[26\]\[22\]](#) [\[4\]](#)

Z6 - Preattentive benefits increase with field of view [\[3\]\[25\]\[26\]\[11\]](#)

Z7 - Quantitative assessment requires position or size variation [\[3\]\[26\]](#)

Z8 - Preserve data to graphic dimensionality [\[24\]\[3\]\[26\]](#)

Z9 - Put the most data in the least space [\[24\]\[26\]](#)

Z10 - Remove the extraneous [\[24\]\[26\]](#)

Z11 - Consider Gestalt Laws [\[25\]\[26\]](#) [\[5\]](#)

Z12 - Provide multiple levels of detail [\[24\]\[25\]\[26\]](#)

Z13 - Integrate text wherever relevant [\[24\]\[25\]\[26\]](#)

6. Forsell's heuristics

The third set of heuristics to use was the Forsell's ten heuristics. [6]

F1 - Information coding

Perception of information is directly dependent on the mapping of data elements to visual objects. This should be enhanced by using realistic characteristics/techniques or the use of additional symbols [B5].

F2 - Minimal actions

Concerns workload with respect to the number of actions necessary to accomplish a goal or a task [E7].

F3 - Flexibility

Flexibility is reflected in the number of possible ways of achieving a given goal. It refers to the means available to customization in order to take into account working strategies, habits and task requirements [E11].

F4 - Orientation and help

Functions like support to control levels of details, redo/undo of actions and representing additional information [B7].

F5 - Spatial organization

Concerns users' orientation in the information space, the distribution of elements in the layout, precision and legibility, efficiency in space usage and distortion of visual elements [B3].

F6 - Consistency

Refers to the way design choices are maintained in similar contexts, and are different when applied to different contexts [E16].

F7 - Recognition rather than recall

The user should not have to memorize a lot of information to carry out tasks [C6].

F8 - Prompting

Refers to all means that help to know all alternatives when several actions are possible depending on the contexts [E1].

F9 - Remove the extraneous

Concerns whether any extra information can be a distraction and take the eye away from seeing the data or making comparisons [D10].

F10 - Data set reduction

Concerns provided features for reducing a data set, their efficiency and ease of use [B9].

Heuristic Evaluation

7. Case Study

The heuristics evaluation was to be performed on Tiago Brito master thesis *Migration Flow Visualization*. The dataset is the result of 3 years (2012, 2013 and 2014) of Portuguese student's applications to college schools. The dataset has 115636 student's applications from 20 districts to 305 schools institutions. More information can be found in Tiago Brito *Migration Flow Visualization* in Information System internet site: [\[7\]](#)

From these data source Tiago Brito created three visualizations: Adjacency Matrix, Chord and Map as we can see in Illustration 1.

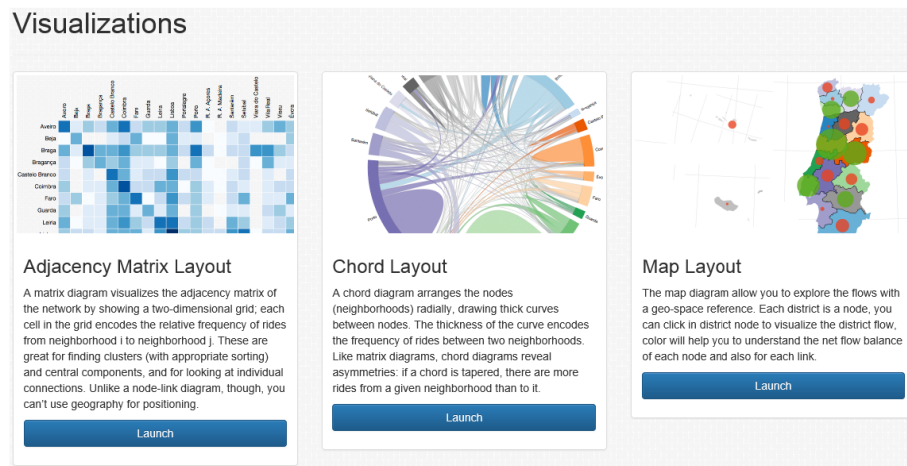


ILLUSTRATION 1 – Visualizations overview on the main site page

8. Adjacency Matrix Visualization

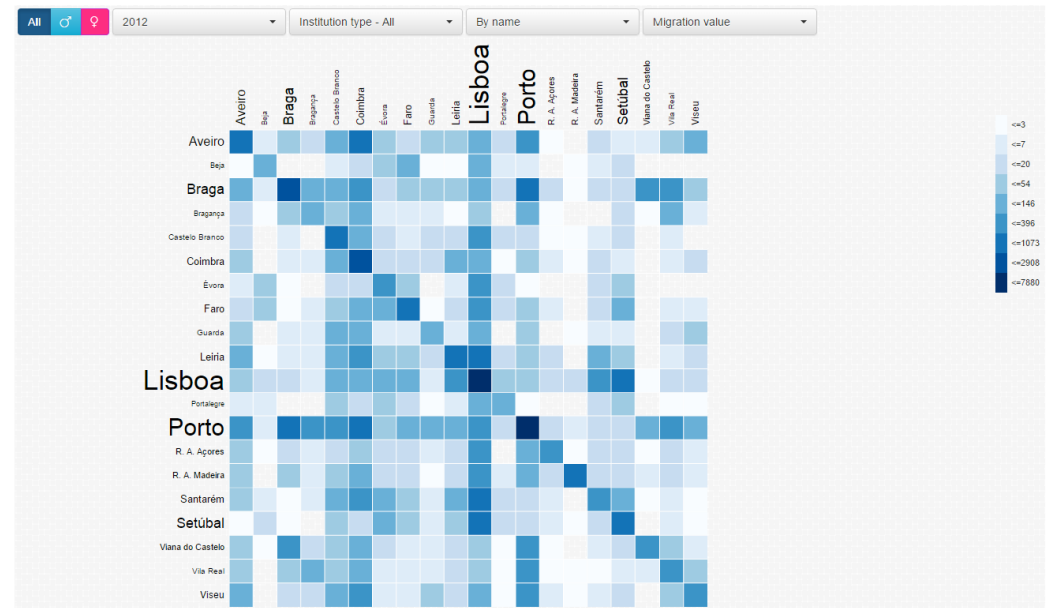


ILLUSTRATION 2 – Adjacency Matrix Visualization

8.1 Nielsen's heuristic evaluation

Heuristic	Applicability	Severity	Comments
N1	4	3	In the All, male and female button it's not perceptible the current selection.
N2	4	3	Concepts not familiar to the user: "Net Flow Balance" and "by nuts2".
N3	4	positive	Home and other visualization links always available on the web page header.
N4	4	2	In the Year selection there is no consolidate view like in "Institution Type – ALL"
N5	0	positive	No errors found.
N6	4	positive	The systems use is perceptible and easy to retrieve information.
N7	4	positive	Easy to interact for both inexperience and experience users.
N8	4	positive	Minimalist design, every unit of information is relevant.
N9	0	positive	No errors found.
N10	4	3	No help section identified. Tooltips should have help information instead of the same as the select box.

Heuristic Evaluation

8.3 Forsell's heuristic evaluation

Heuristic	Applicability	Severity	Comments
F1	4	Positive	Information perception is directly dependent on mapping of data elements.
F2	4	Positive	Achieving a goal is typically achieved by one or two clicks.
F3	3	Positive	There is only one way to achieve a given goal.
F4	4	Positive	Levels of details on mouse over (tooltip) and on mouse click. Home and other visualization links always available on the web page header.
F5	4	Positive	Layout is precise, legible and efficient in space usage.
F6	4	Positive	Design choices are maintained.
F7	4	Positive	The systems use is perceptible and easy to retrieve information.
F8	4	Positive	Alternatives are in select boxes.
F9	4	Positive	The visualization is clean and simple.
F10	4	Positive	Data set reduction is achieved by using the select boxes.

Heuristic Evaluation

8.2 Zuk's heuristic evaluation

Heuristic	Applicability	Severity	Comments
Z1	3	Positive	Visual variable has length.
Z2	4	Positive	There is no reading order from color, however there is value order.
Z3	4	Positive	Matrix square size is the same independently of the color.
Z4	4	3	Background color should help distinguish more the matrix.
Z5	4	Positive	However different, does not change interpretation for color blind persons. See Illustration 3.
Z6	4	Positive	It's preattentive that Lisboa and Porto lines and columns are darker.
Z7	4	Positive	City label are sized by population, is perceptible that the bigger the more value it represents. It's possible to order by population.
Z8	4	Positive	Data to graphic dimensionality is preserved.
Z9	4	Positive	Data space relation is adequate.
Z10	4	Positive	The visualization is clean and simple.
Z11	4	Positive	No principal of grouping are seen in the visualization. However Lisboa and Port lines and columns are darker. City label next to respective lines and columns.
Z12	4	Positive	Multiple levels of detail are available on mouse over (tooltip) and on mouse click.
Z13	4	Positive	Text exist to support data visualization, on mouse over (tooltip) and on mouse click.

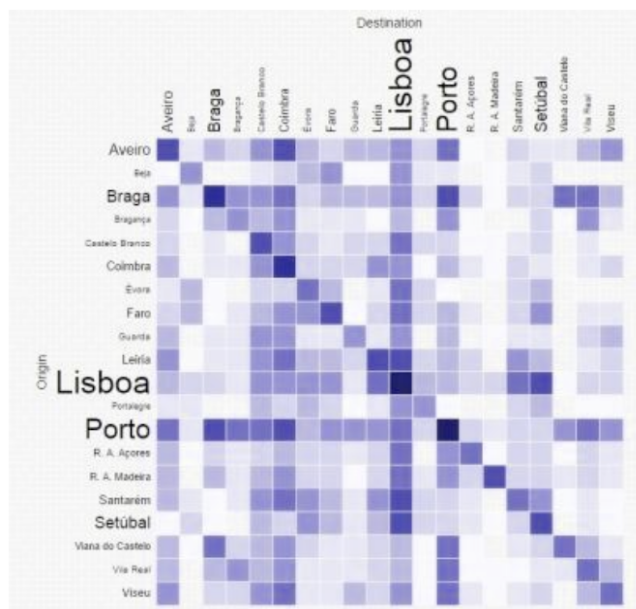


ILLUSTRATION 3 – Adjacency Matrix Visualization with Protanopia Filter

Heuristic Evaluation

9. Chord Visualization

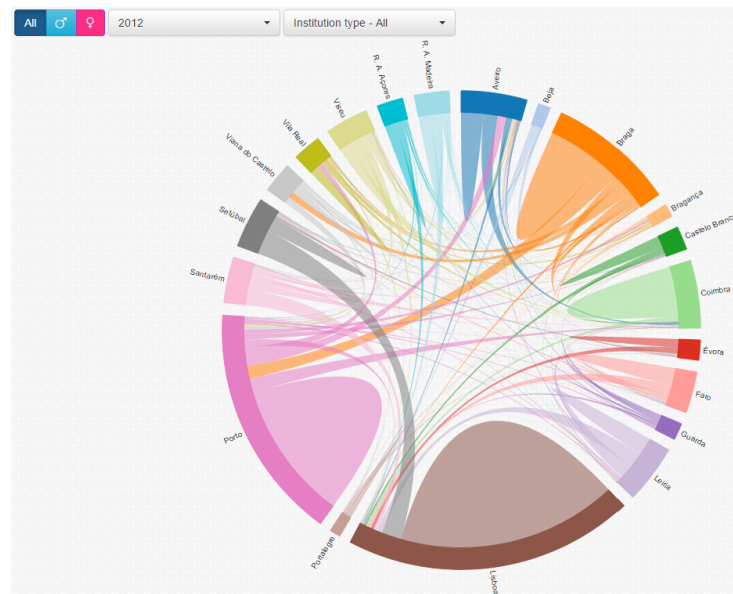


ILLUSTRATION 4 – Chord Visualization

9.1 Nielsen's heuristic evaluation

Heuristic	Applicability	Severity	Comments
N1	4	3	In the All, male and female button it's not perceptible the current selection.
N2	4	3	No details on mouse over (tooltip), only on mouse click.
N3	4	positive	Home and other visualization links always available on the web page header.
N4	4	2	Not consistent use of "Chord" and "Circular" in the several interfaces. In the Year selection there is no consolidate view like in "Institution Type – ALL".
N5	0	positive	No errors found.
N6	4	positive	The systems use is perceptible and easy to retrieve information.
N7	4	positive	Easy to interact for both inexperience and experience users.
N8	4	positive	Minimalist design, however data values should be available on mouse over.
N9	0	positive	No errors found.
N10	4	3	No help section identified. Tooltips should have help information instead of the same as the select box.

Heuristic Evaluation

9.2 Zuk's heuristic evaluation

Heuristic	Applicability	Severity	Comments
Z1	3	Positive	Visual variable has length.
Z2	4	Positive	There is no reading order from color, the colors only serve to distinct one district from another.
Z3	4	Positive	Greater size represents greater data values.
Z4	4	3	Background color affects perception of some lighter chord colors.
Z5	2	3	Interpretation is different for color blind persons. See Illustration 5.
Z6	4	Positive	It's preattentive that Lisboa and Porto have greater data values.
Z7	4	Positive	Quantitative makes size variation.
Z8	4	Positive	Data to graphic dimensionality is preserved.
Z9	4	Positive	Data space relation is adequate.
Z10	4	Positive	The visualization is clean and simple.
Z11	4	Positive	No principal of grouping are seen in the visualization. However Lisboa and Port areas are a lot superior the other districts.
Z12	4	3	Multiple levels of detail are available on mouse click, however no data values on mouse over (tooltip).
Z13	4	3	No text Tooltips exists to support data visualization, only on mouse click.

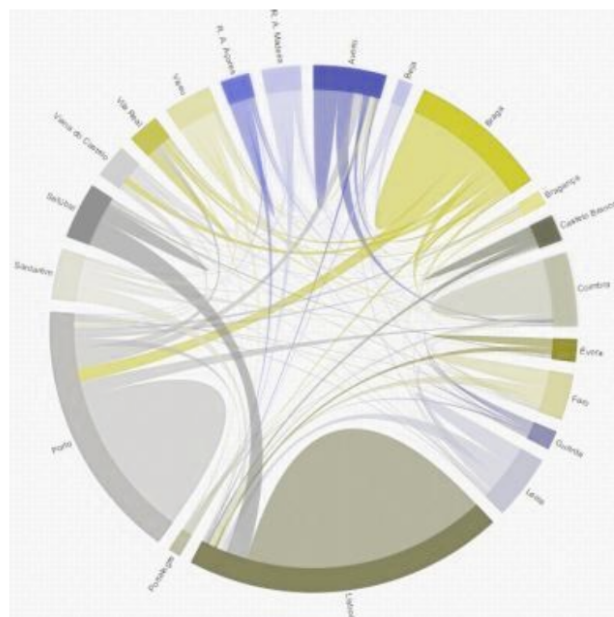


ILLUSTRATION 5 – Chord Visualization with Protanopia Filter

Heuristic Evaluation

9.3 Forsell's heuristic evaluation

Heuristic	Applicability	Severity	Comments
F1	4	Positive	Information perception is directly dependent on mapping of data elements.
F2	4	Positive	Achieving a goal is typically achieved by one or two clicks.
F3	3	Positive	There is only one way to achieve a given goal.
F4	4	3	No levels of detail on mouse over (tooltip), only on mouse click. Home and other visualization links always available on the web page header.
F5	4	Positive	Layout is precise, legible and efficient in space usage.
F6	4	3	Design choices are not maintained, instead of mouse over we need to click the visualization.
F7	4	Positive	The systems use is perceptible and easy to retrieve information.
F8	4	Positive	Alternatives are in select boxes and on mouse click.
F9	4	Positive	The visualization is clean and simple.
F10	4	Positive	Data set reduction is achieved by using the select boxes.

Heuristic Evaluation

10. Map Visualization

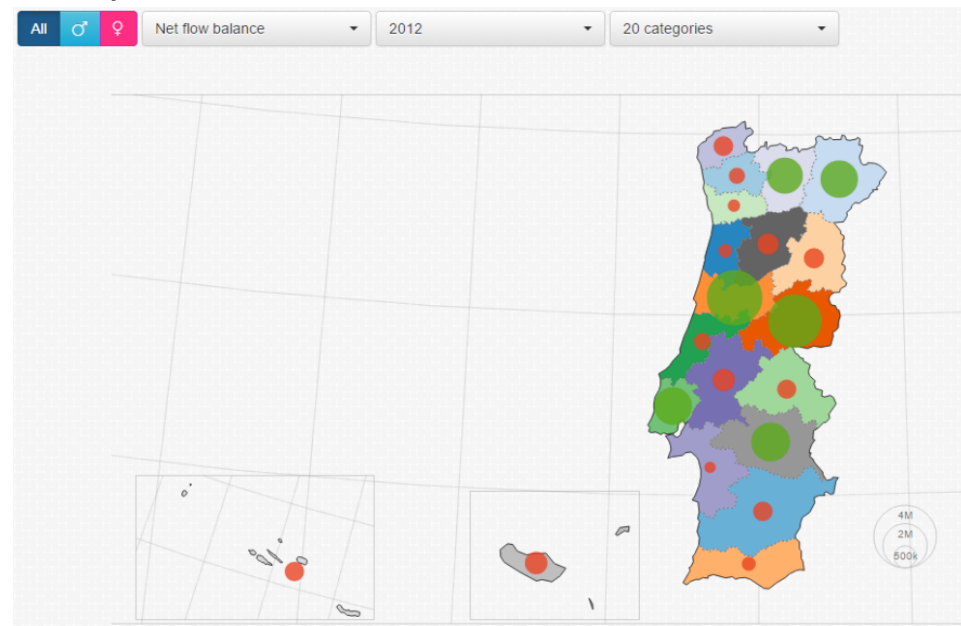


ILLUSTRATION 6 – Map Visualization

10.1 Nielsen's heuristic evaluation

Heuristic	Applicability	Severity	Comments
N1	4	3	In the All, male and female button it's not perceptible the current selection.
N2	4	3	Concepts not familiar to the user: "Population Choropleth".
N3	4	positive	Home and other visualization links always available on the web page header.
N4	4	2	In the Year selection there is no consolidate view like in "Institution Type – ALL".
N5	0	positive	No errors found.
N6	4	positive	The systems use is perceptible and easy to retrieve information.
N7	4	positive	Easy to interact for both inexperience and experience users.
N8	4	positive	Minimalist design, every unit of information is relevant.
N9	0	positive	No errors found.
N10	4	3	No help section identified. Tooltips should have help information instead of the same as the select box.

Heuristic Evaluation

10.2 Zuk's heuristic evaluation

Heuristic	Applicability	Severity	Comments
Z1	3	Positive	Visual variable has length.
Z2	4	Positive	There is no reading order from colors, colors indicates positive or negative flow.
Z3	4	Positive	Greater size represents greater data values.
Z4	4	Positive	Background color does not affects perception.
Z5	4	3	Interpretation is different for color blind persons. See Illustration 7.
Z6	4	Positive	It's preattentive that Coimbra and Castelo Branco have superior data values in the default selection.
Z7	4	Positive	Quantitative makes size variation, color positive/negative values.
Z8	4	Positive	Data to graphic dimensionality is preserved.
Z9	4	Positive	Data space relation is adequate.
Z10	4	Positive	The visualization is clean and simple.
Z11	4	Positive	No principal of grouping are seen in the visualization. However Coimbra and Castelo Branco areas have greater perception then other districts.
Z12	4	Positive	Multiple levels of detail are available on mouse over and on mouse click.
Z13	4	Positive	Text exist to support data visualization, on mouse over (tooltip) and on mouse click.

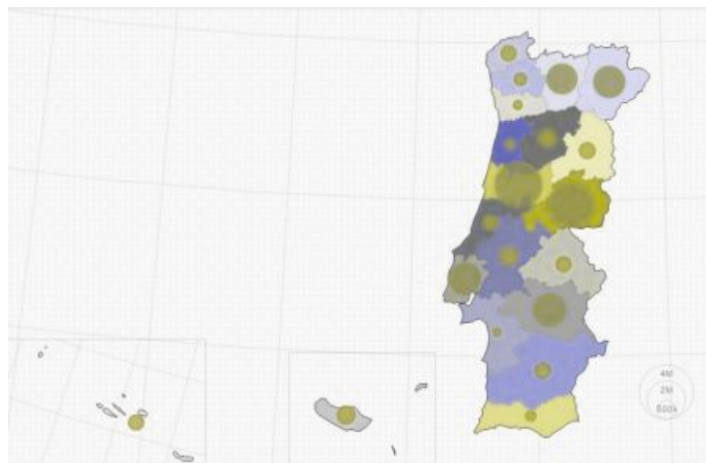


ILLUSTRATION 7 – Map Visualization with Protanopia Filter

Heuristic Evaluation

10.3 Forsell's heuristic evaluation

Heuristic	Applicability	Severity	Comments
F1	4	Positive	Information perception is directly dependent on mapping of data elements.
F2	4	Positive	Achieving a goal is typically achieved by one or two clicks.
F3	3	Positive	There is only one way to achieve a given goal.
F4	4	3	Details on mouse over (tooltip) and on mouse click. Home and other visualization links always available on the web page header.
F5	4	Positive	Layout is precise, legible and efficient in space usage.
F6	4	3	Design choices are maintained and there is that instruction on the summary.
F7	4	Positive	The systems use is perceptible and easy to retrieve information.
F8	4	Positive	Alternatives are in select boxes and on mouse click.
F9	4	Positive	The visualization is clean and simple.
F10	4	Positive	Data set reduction is achieved by using the select boxes.



11. Conclusions

The objective of this report was to analyze literature on Information Visualization, use a set of heuristics and evaluate its applicability to this specific case study.

The results of this test show that heuristics are relevant and applicable to Information Visualization. However in my opinion the results also suggest that not all heuristics are relevant to the system, contradicting the generic objective of heuristics.

It was possible to verify some similar heuristics in the three heuristics sets like Help, recognition, Consistency/standardization and minimalistic/remove the extraneous.

The Visualizations shared the same web site and consequently the same page header, however the visualizations and option were different from each other. Besides the perception of information that was different depending on the user intention, for instance Matrix visualization was better for exact values, Chord visualization was better for flow perception and Map for geographic perception. Some display information were achieved from different interaction, for example on mouse over only in the matrix visualization and more information on mouse click in the two others visualizations.

As future work, I suggest the extension of the heuristics set and number of evaluators.

12. References

- [1] <http://www.aholzinger.at/publications/>
- [2] <http://www.nngroup.com/articles/ten-usability-heuristics/>
- [3] Heuristics for Information Visualization Evaluation [ZUK 2006]
- [4] <http://www.color-blindness.com/coblis/coblis.html>
- [5] http://en.wikipedia.org/wiki/Principles_of_grouping
- [6] An Heuristic Set for Evaluation in Information Visualization [FORSELL 2010]
- [7] <http://www.portugal-migration.info/>

Avaliação de Ana Gonçalves

UNIVERSIDADE DE AVEIRO
DEPARTAMENTO DE ELETRÓNICA, TELECOMUNICAÇÕES E INFORMÁTICA

Relatório de Avaliação de Usabilidade

ANA GONÇALVES
23031

18 de Maio de 2015

Mestrado em Sistemas de Informação
Visualização da Informação
Beatriz Sousa Santos

1 Introdução e sobre a avaliação de usabilidade

Este trabalho surge no contexto da disciplina Visualização da Informação do primeiro ano do segundo ciclo em Sistemas de Informação, lecionado no Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro. Foi proposta a avaliação da usabilidade de uma aplicação para a visualização e análise de fluxos de migração de estudantes do ensino superior, em Portugal.

1.1 Classificar um problema de usabilidade

Segundo Nielsen [1] a severidade de um problema de usabilidade resulta da combinação de três fatores:

- a frequência com que o problema ocorre: ocorre muitas vezes ou é raro acontecer?
- o impacto gerado se o problema ocorrer: se o utilizador se deparar com este problema consegue ultrapassá-lo? com maior ou menor dificuldade?
- a persistência do problema: é um problema que uma vez conhecido pelo utilizador, este sabe como o evitar ou é um problema que incomoda repetidamente o utilizador?

Com base na severidade é possível pontuar um problema de usabilidade, o que permite também priorizar a sua resolução. A escala que se apresenta em seguida foi a escala adotada.

- 0: não concordo que este seja um problema de usabilidade
- 1: é meramente um problema de cosmética, não precisa ser corrigido a não ser que haja disponibilidade
- 2: problema de usabilidade menor, baixa prioridade na resolução deste problema
- 3: problema de usabilidade grave, é importante corrigi-lo, prioridade elevada na sua resolução
- 4: problema de usabilidade crítico/bloqueante, problema terá que estar resolvido antes que o produto possa ser distribuído

1.2 Processo de avaliação - visão geral

Durante a sessão de avaliação [2] o avaliador navega na aplicação analisando várias vezes diferentes elementos/componentes, procurando identificar problemas de usabilidade com base numa lista de princípios - as heurísticas.

Recomenda-se que o avaliador percorra a interface pelo menos duas vezes. A primeira tem como objetivo obter uma noção da forma de navegação/interação com a interface e o âmbito da aplicação a avaliar. A segunda vez permite ao avaliador focar-se em elementos específicos, conhecendo já o seu contexto na aplicação.

O resultado final da avaliação é uma lista de problemas de usabilidade devidamente referenciados com os princípios de usabilidade que, na opinião do avaliador, não são satisfeitos.

Para a avaliação aqui apresentada, foram previamente estudados dois conjuntos de heurísticas. Estes dois conjuntos são apresentados nas duas sub secções seguintes.

1.3 Dois conjuntos de heurísticas

Segundo Zuk et al. [3], o conjunto de heurísticas apresentado em [6] é mais apropriado para avaliar um aplicação do ponto de vista da percepção. Numa outra abordagem, Forsell e Johansson [4] procuraram identificar um conjunto de dez heurísticas que permitissem cobrir o maior número de problemas de usabilidade de entre um conjunto de exemplos devidamente identificados. As duas sub secções seguintes apresentam, respetivamente, o conjunto de heurísticas apresentado em [6] - designado por Conjunto de heurísticas de Zuk - e o conjunto de heurísticas apresentado em [4] - designado por Conjunto de heurísticas de Forsell.

1.3.1 Conjunto de heurísticas de Zuk

Garantir que os elementos visuais têm dimensão suficiente - *Ensure visual variable has sufficient length*

A cor não permite definir/transmitir qualquer tipo de ordem - *Don't expect a reading order from color*

Percepção de cores varia com o tamanho do item colorido - *Color perception varies with size of colored item*

Conforme o ambiente/contexto a percepção da cor varia - *Local contrast affects color & gray perception*

Ter em consideração pessoas daltónicas - *Consider people with color blindness*

A percepção subconsciente melhora com o aumento do campo de visão - *Preattentive benefits increase with field of view*

Avaliação quantitativa requer variação do tamanho ou posição - *Quantitative assessment requires position or size variation*

Preservar a proporção dos dados para dimensão dos gráficos - *Preserve data to graphic dimensionality*

Relacionado com a densidade dos dados - *Put the most data in the least space*

Não existência de informação adicional que distraia o utilizador do que é relevante ou dificulte a realização de comparações - *Remove the extraneous (ink)*

Utilização de propriedades como proximidade, similaridade, continuidade, simetria, fecho, conexão e tamanho relativo de modo a facilitar a percepção de padrões - *Consider Gestalt Laws*

O utilizador tem controlo sobre o nível de detalhe - *Provide multiple levels of detail*

Inclusão de texto sempre que possa ser útil; *tooltips*, por exemplo - *Integrate text wherever relevant*

1.3.2 Conjunto de heurísticas de Forsell

B5: Codificação da informação - *Information coding*

Além do mapeamento dos dados em elementos visuais, a utilização de símbolos ou técnicas realistas ajuda na percepção/compreensão dos elementos de informação. Por exemplo, grupos de elementos [7].

E7: Número mínimo de passos - *Minimal actions*

É desejável que o número de passos a executar para atingir um objetivo seja o menor possível [5].

E11: Flexibilidade - *Flexibility*

Flexibilidade é a capacidade de a interface se adaptar às necessidades do utilizador oferecendo, por exemplo, diferentes formas/caminhos de atingir um objetivo ou a possibilidade de personalização da própria interface [5].

B7: Orientação e ajuda - *Orientation and help*

Engloba funcionalidades como controlo do nível de detalhe, desfazer/refazer ações e representação de informação adicional (por exemplo o caminho que o utilizador seguiu na navegação numa estrutura complexa) [7].

B3: Organização espacial - *Spatial organization*

Está relacionado com o *layout* geral de uma representação visual: se é fácil encontrar um determinado elemento de informação, se é fácil ter consciência/percecionar a distribuição dos elementos visuais na representação [7].

E16: Consistência - *Consistency*

Uma interface é consistente quando aplica o mesmo tipo de design/solução em situações similares, quando o *look&feel* é o mesmo nos vários ecrãs [5].

C6: Reconhecimento, ao invés da memória - *Recognition rather than recall*

Não deve ser necessário o utilizador memorizar como executar as tarefas na aplicação. Instruções de como usar a aplicação devem estar visíveis e acessíveis sempre que necessário [8].

E1: Solicitações e mensagens - *Prompting*

Consiste na ajuda disponibilizada e que permite ao utilizador conhecer as alternativas quando várias ações são possíveis. Exemplos: informação sobre o estado ou contexto da aplicação, ajuda contextualizada, dicas sobre o preenchimento de campos [5].

D10: Remover o não essencial - *Remove the extraneous*

Não existência de informação adicional que distraia o utilizador do que é relevante ou dificulte a realização de comparações. Corresponde ao princípio *Data-Ink maximization* de Tufte e que também faz parte do conjunto de heurísticas de Zuk [6].

B9: Redução do conjunto de dados - *Data set reduction*

Se existem funcionalidades que permitem reduzir o conjunto de dados: filtragem, agrupamento e simplificação removendo informação irrelevante [7].

1.3.3 A escolha

Neste trabalho, os fatores que determinam a escolha do conjunto de heurísticas a usar são o grau de confiança que o avaliador tem na correta interpretação dos princípios e a adequação dos mesmos ao objetivo da avaliação. Na minha opinião, o conjunto de heurísticas de Forsell é de mais fácil interpretação, sendo o de Zuk mais focado na avaliação da representação visual. No conjunto de Forsell a avaliação da representação e da usabilidade estão a par.

O conjunto de heurísticas escolhido para servir de referência na avaliação foi o conjunto de Forsell. Foram duas as razões que motivaram esta opção:

- este conjunto de princípios permite, de forma relativamente equilibrada, avaliar do ponto de vista de perceção e do ponto de vista da usabilidade;
- considero ter compreendido melhor este conjunto de heurísticas.

2 Avaliação de usabilidade

2.1 Aspetos que estão de acordo com as heurísticas

1. Na visualização *Matrix*, quando se coloca o ponteiro do rato sobre um quadrado, ambas as etiquetas dos distritos relacionados mudam de cor, facilitando a leitura da matriz. Ver o exemplo da figura 1. - Heurísticas B5, B7.
2. Na visualização *Matrix*, tanto a informação que é apresentada quando se coloca o ponteiro do rato sobre um quadrado como a que é mostrada quando se clica sobre o quadrado, correspondem à disponibilização de mais detalhe. - Heurística B7.
3. As visualizações *Matrix* e *Map* através da opção de cor, oferecem alguma flexibilidade ao utilizador na visualização. - Heurística E11.
4. Na visualização *Chord*, há algum controlo do nível de detalhe quando se coloca o ponteiro do rato sobre um distrito ou apenas uma corda, sendo mostrada informação relativa apenas àquele distrito ou fluxo. - Heurística B7.
5. Na visualização *Map*, o uso do mapa de Portugal como base, permite analisar os fluxos migratórios com o contexto da adjacência (ou não) entre distritos. - Heurística B5.
6. Na visualização *Map*, a informação sobre os fluxos que se obtém ao clicar sobre o círculo representativo de um determinado distrito, também corresponde à disponibilização de mais detalhe. - Heurística B7.
7. Nas várias visualizações, o uso de símbolos para representar o género e a sua utilização em botões com funcionalidade tipo *radio button* permite uma filtragem rápida, não ocupando muito espaço. Ver o exemplo da figura 1. - Heurística B5.
8. É disponibilizada ao utilizador uma variedade adequada de filtros de fácil utilização, consistentes nas diferentes visualizações. - Heurísticas B9, E7 e E11.
9. O uso de caixas de seleção para os filtros, uma vez que são uma abordagem comum, facilitam a compreensão da funcionalidade. - Heurística C6.
10. Todas as visualizações têm hiperligações diretas através das opções no cabeçalho. - Heurística E7.

11. A breve descrição da visualização com dicas de utilização ajudam a colmatar a falta de ajuda. - Heurística E1.
12. De um modo geral, não há informação a mais que distraia o utilizador. - Heurística D10.

2.2 Aspetos que poderiam ser melhorados

2.2.1 Gravidade 1

1. Nas visualizações *Matrix* e *Map* as legendas não estão localizadas na mesma zona da página. Esta situação é ilustrada na figura 2. Em particular, na visualização *Map*, quando selecionado em *Map color* o valor *Population choropleth*, não é apresentada legenda. - Heurísticas B7, B3 e E16.
2. Na página inicial, o texto que antecede (e porque o tamanho de letra é grande) distrai das visualizações. - Heurística D10.
3. O *Help* não funciona. - Heurística E1.
4. Na visualização *Matrix*, qual é a lógica que determina o tamanho de letra para os diferentes distritos? Por exemplo, Aveiro tem letra tamanho 16 enquanto Coimbra tem letra tamanho 12. No entanto, aparentemente, Coimbra apresenta valores maiores em cada um dos parâmetros. - Heurística E16.

2.2.2 Gravidade 2

1. Na visualização *Matrix*, uma vez realizado o clique sobre um quadrado, não se consegue fechar a listagem de instituições. - Heurísticas B7 e E1.
2. Na visualização *Chord* seria interessante conseguir fixar as cordas para um determinado distrito. Por exemplo, se passar o ponteiro do rato sobre o distrito de Braga, apenas os fluxos relativos a Braga são visíveis; se então eu quiser obter mais detalhe sobre algum desses fluxos, não será uma tarefa fácil, pois ao mover o rato, o conjunto de cordas visíveis altera-se. - Heurística B7.
3. Na visualização *Map*, por exemplo quando se seleciona o distrito de Lisboa, não são visíveis os fluxos de saída, tal como ilustrado na figura 3. - Heurística B3.
4. Habitualmente, a zona de um site/aplicação web com o nome do mesmo, se clicável, leva o utilizador para a página inicial. No caso da aplicação avaliada, a zona no cabeçalho da página com o texto "Migration Flow Visualization", apesar de clicável, não tem qualquer tipo de evento associado. Na minha experiência, só ao fim de algum tempo de utilização é que me apercebi que existia uma opção "Home". - Heurística C6.
5. Nas várias visualizações, tendo algum tipo de filtragem, se se fizer *refresh* no navegador, apesar de a informação apresentada ser relativa aos valores de filtragem por omissão, o que aparece nas caixas de seleção é a filtragem anterior, induzindo o utilizador em erro. - Heurística E1.
6. Na visualização *Matrix*, os valores apresentados para o *net flow balance* parecem ser calculados subtraindo o número de estudantes que sai ao número de estudantes que chega a um determinado distrito. No entanto, na descrição do *net flow balance* é indicado o contrário. - Heurística E1.

2.2.3 Gravidade 3

1. Na visualização *Matrix* a informação adicional quando se clica sobre um quadrado rapidamente perde o seu contexto, bastando uma pequena distração do utilizador. Por exemplo, na figura 4, pelo nome das instituições listadas, infere-se que o distrito de destino é a Região Autónoma dos Açores, mas não se sabe qual o distrito de origem. - Heurística B7.

2.2.4 Gravidade 4

1. Na visualização *Chord*, a informação adicional por vezes interfere com a visualização do gráfico, tal como ilustrado na figura 5, não sendo possível removê-la. - Heurísticas B3 e E11.
2. Na visualização *Chord*, quando se filtram os dados, apesar de graficamente haver alterações, a informação adicional disponibilizada é a mesma. Esta situação é demonstrada nas figuras 6 e 7. - Heurística E1.

2.3 Outras considerações

- Falta uma versão em português, uma vez que os dados são sobre estudantes portugueses em Portugal.
- Num ecrã de menores dimensões, a apresentação das várias visualizações (presente na página inicial) não é completamente visível.
- Na visualização *Matrix*, a legenda e informação visualizada parecem não estar coerentes. No exemplo da figura 8, o quadrado sobre o qual está o ponteiro do rato apresenta a cor roxa, que na legenda é identificado como sendo *None*, no entanto, o que é mostrado na informação adicional é que o *net flow balance* é 12.
- Nas visualizações *Matrix* e *Chord*, quando se seleciona "Ensino Superior Público Universitário" ou "Ensino Superior Público Politécnico" na caixa "Institution type", o texto aparece cortado. Não se justifica, uma vez que há espaço.

3 Conclusões

Considero que, de um modo geral, houve preocupação com o respeito pelas boas práticas. A maioria dos problemas encontrados são menores, sendo que os problemas mais graves estão relacionados com a informação que é mostrada (que nem sempre está de acordo com filtros aplicados) ou com situações que apenas ocorrem quando se posiciona o rato num determinado ponto da visualização.

4 Anexo - Imagens

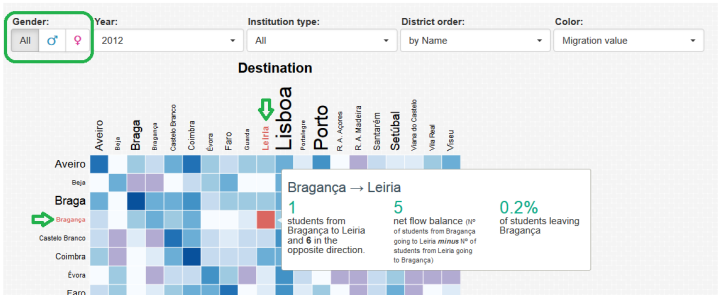


Figura 1: Visualização *Matrix*.

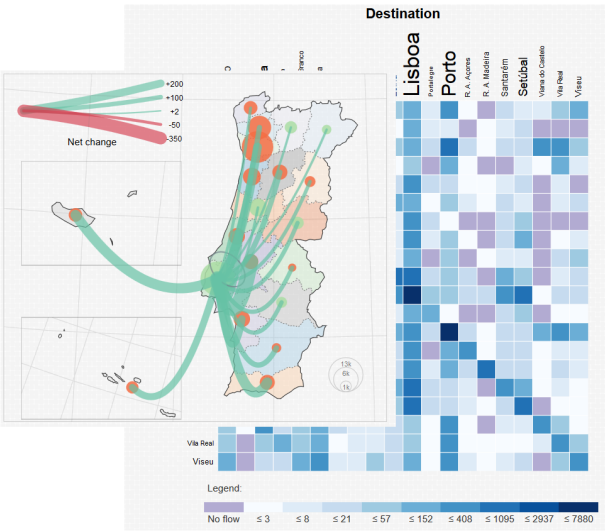
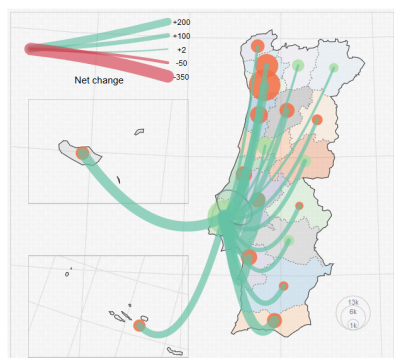
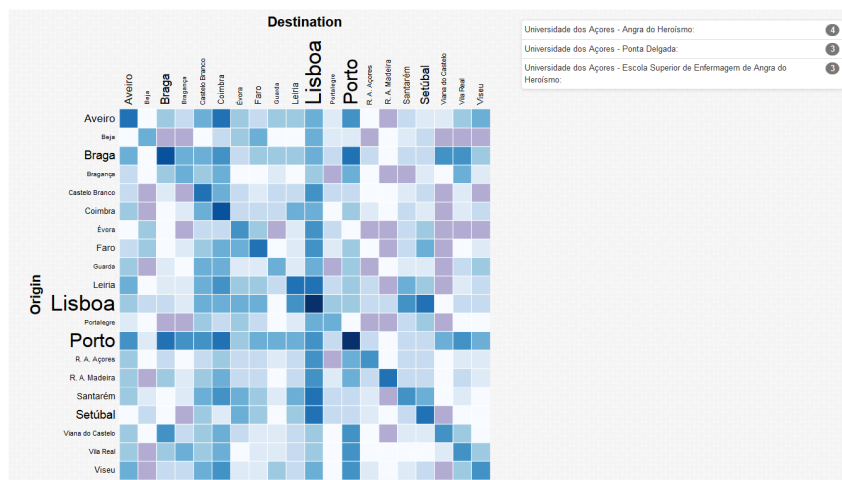


Figura 2: Visualizações *Matrix* e *Map* - legendas distintas.

Figura 3: Visualização *Map* - informação adicional.Figura 4: Visualização *Matrix* - informação adicional.

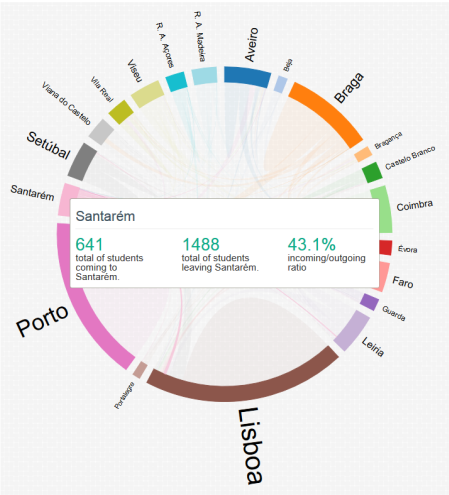


Figura 5: Visualização *Chord* - informação adicional esconde gráfico.

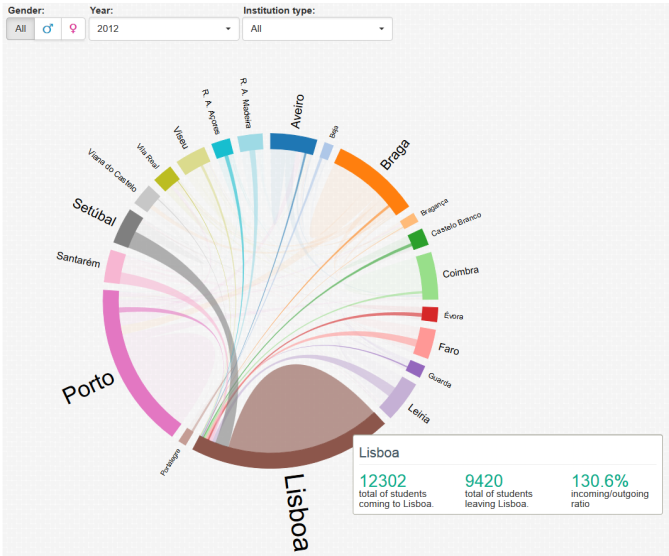
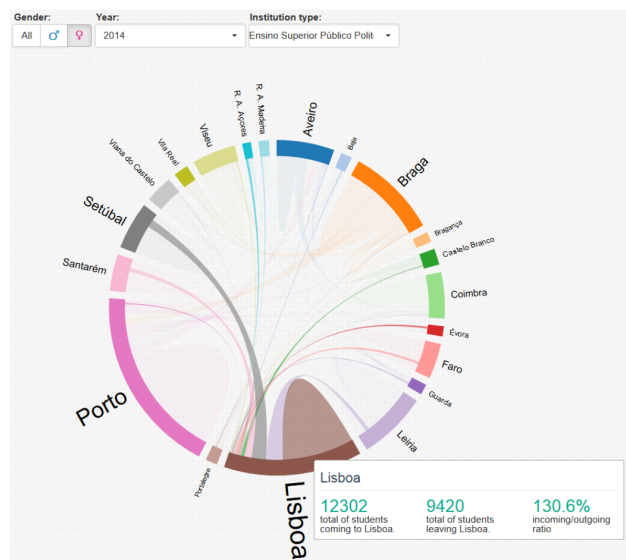
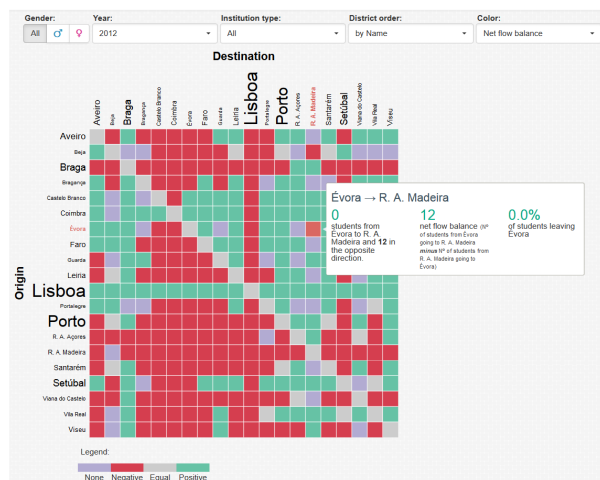


Figura 6: Visualização *Chord* - sem filtros.

Figura 7: Visualização *Chord* - com filtros.Figura 8: Visualização *Matrix* - net flow balance.

Referências

- [1] Nielsen, J. 1995. Severity Ratings for Usability Problems.
<http://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/>
- [2] Nielsen, J. 1995. How to Conduct a Heuristic Evaluation.
<http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/>
- [3] Zuk, T., Schlesier, L., Neumann, P., Hancock, M.S. and Carpendale, S. 2006. Heuristics for Information Visualization Evaluation. In *Proc BELIV'06* ACM Press 1-6.
- [4] Forsell, C. and Johansson, J. 2010. An Heuristic Set for Evaluation in Information Visualization. In *Proc AVI'10* ACM Press 199-206.
- [5] Scapin, D.L. and , Bastien, J.M.C. 1997. Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & Information Technology* 16 (4-5) 220-231.
- [6] Zuk, T. and Carpendale, S. 2006. Theoretical analysis of uncertainty visualizations. In *Proc SPIE-IS&T Electronic Imaging*, SPIE, vol. 6060, 606007.
- [7] Freitas, C.M.D.S., Luzzardi, P.R.G., Cava, R.A., Winckler, M.A.A., Pimenta, M.S. and Nedel, L.P. 2002. Evaluating Usability of Information Visualization Techniques. In *Proc 5th Workshop on Human Factors in Computer Systems (IHC)* Brazilian Computer Society Press 40-51.
- [8] Nielsen, J. 1995. 10 Usability Heuristics for User Interface Design.
<http://www.nngroup.com/articles/ten-usability-heuristics/>

